Talanta 194 (2019) 563–575

Contents lists available at ScienceDirect

Talanta

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

A chemometric strategy for forensic analysis of condom residues: Identification and marker profiling of condom brands from direct analysis in real time-high resolution mass spectrometric chemical signatures

Allix M. Coon, Samira Beyramysoltan, Rabi A. Musah*

Department of Chemistry, State University of New York at Albany, 1400 Washington Ave, Albany, NY 12222, United States

ARTICLE INFO

Keywords:
Sexual assault
Condom brand
Lubricant
Direct analysis in real time mass spectrometry
Brand
Chemometrics
Conformal prediction

ABSTRACT

The rapid and accurate identification of condom-derived lubricant traces takes on heightened importance in sexual assault cases where the assailant has used a condom in order to avoid leaving behind incriminating DNA evidence. Previous reports have demonstrated that a variety of techniques can be used to confirm that a given residue is condom-derived, based on the detection of spermicides, slip agents and/or other common additives. However, limited success has been achieved in differentiating brands from among a broad range of condom types. In this study, the utility of direct analysis in real time-high resolution mass spectrometry (DART-HRMS) combined with chemometrics, for the rapid and accurate attribution of brands to condom residues of various types, was explored and developed. A database of condom residue spectra comprised of 110 different condom types representing 16 brands was generated, with the spectra serving as representative fingerprints for each brand. The spectral fingerprints were subjected to pre-processing prior to the application of Partial Least Squares-Discriminant Analysis (PLS-DA) which was used to generate a classifier that permitted identification of condom brands with an accuracy of 97.4%. An additional criterion was imposed on the PLS-DA to provide the confidence level and credibility of each prediction. The effect of time since deposition, the presence of contaminants and the influence of residue transfer on the prediction accuracy of the model were also assessed.

The results from Sparse Discriminant Analysis (SDA) and PLS-DA were followed by application of the Student's *t*-test to determine *m/z* values representative of small-molecule markers that were most important for defining brand classes. The *m/z* values revealed by the two methods were found to be consistent in indicating which masses were representative of markers. The SDA method also provided low-dimensional views of the discriminative directions for classification of condom residues, thereby enabling easy visualization of the relationship between the indicated *m/z* values and brand discrimination. The results further revealed a subset of 14 *m/z* values that were observed in all 110 condoms representing the 16 brands, and some of these may serve as potential universal small-molecule condom markers. Overall, the results show that the DART-HRMS database of condom residue spectra can be used to identify residues based on differences in chemical components peculiar to each brand. The database can be readily expanded to include more condoms.

1. Introduction

The ability to identify condom-derived trace evidence is gaining in importance because of the increasing frequency with which perpetrators of sexual assault use condoms in order to avoid leaving behind incriminating DNA evidence. When DNA is absent because a condom was used, the condom-derived lubricant traces left behind may have significant evidentiary value. In such cases, it would be highly beneficial to be able to not only confirm the presence of the lubricant, but perhaps even identify the brand of condom used. Blackledge and Vincenti reported two cases in which condom traces provided important associative evidence. In both instances, confirmation of the use of a condom was based on detection of the silicone-based lubricant polydimethylsiloxane (PDMS) by FT-IR and desorption chemical

Abbreviations: DCI, desorption chemical ionization; DNA, deoxyribonucleic acid; FT-IR, Fourier Transform Infrared Spectroscopy; GC-MS, Gas Chromatography Mass Spectrometry; MALDI-MS, Matrix Assisted Laser Desorption Ionization Mass Spectrometry; LC-MS, Liquid Chromatography Mass Spectrometry; DRIFTS, Diffuse Reflectance Infrared Fourier Transform Spectroscopy; NMR, Nuclear Magnetic Resonance

* Corresponding author.
E-mail address: rmusah@albany.edu (R.A. Musah).

https://doi.org/10.1016/j.talanta.2018.09.101
Received 16 July 2018; Received in revised form 25 September 2018; Accepted 27 September 2018
Available online 27 September 2018
0039-9140/

ionization (DCI)-MS [1,2]. The spermicidal detergent nonoxynol-9 (N-9) was also detected by FT-IR [1]. Other early studies have focused on alternative methods for the identification of PDMS and/or identification of specific compounds commonly found in condoms, such as polyethylene glycol (PEG) and N-9. For example, Campbell et al. showed that PDMS and PEG could be identified in condom residue using conventional GC-MS and pyrolysis GC-MS [3]. Further studies also expanded on the types of analytical instruments used to identify condom residues. Examples include the application of Raman spectroscopy and chemical imaging to the detection of PDMS [4–6], MALDI-MS imaging, attenuated total reflectance (ATR)-FTIR of condom lubricant-contaminated fingermarks for the detection of N-9, the spermicide octoxynol-9 and PEG [6] and the use of capillary electrophoresis [7].

In a comprehensive study in which several methods for distinguishing between condom types were assessed (i.e. fluorescence microscopy/spectroscopy, ATR- and diffuse reflectance-IR spectroscopy, GC-MS, LC-MS and pyrolysis GC-MS), Maynard et al. analyzed 58 condom types representing 10 brands. The authors developed a workflow for rape kit analysis based on whether samples were extracted in hexane or methanol (MeOH), and using a variety of instrumental techniques (i.e. DRIFTS, pyrolysis GC-MS and GC-MS for hexane samples; and DRIFTS and LC-MS for MeOH samples). The approach enabled differentiation of petrolatum, PDMS, PEG and N-9 containing lubricants, and could be used to uniquely identify 11 condom samples [8]. Lee et al. used solid state and solution ^{1}H NMR in a study of 38 condoms representing 12 different manufacturers. While this approach enabled discrimination between condoms based on differences in the composition of the rubber of which the condoms were comprised, solution ^{1}H NMR furnished information on the distinguishing features associated with the presence of PEG, PDMS and N-9. By this method, 15 of the 38 condoms assayed could be individualized [9]. Hollenbeck et al. demonstrated that trace amounts of N-9 could be recovered and identified in vaginal swabs collected post-coitus by reverse phase liquid chromatography/electrospray ionization mass spectrometry, nano electrospray mass spectrometry and high resolution MALDI Fourier Transform mass spectrometry [10]. Musah et al. developed a GC-MS protocol for analysis of N-9 [11]. They subsequently demonstrated that an efficient way in which to circumvent the time-consuming sample preparation steps associated with analyzing condom residues by conventional methods is to use the ambient ionization mass spectrometric technique direct analysis in real time-high resolution mass spectrometry (DART-HRMS) [12]. DART-HRMS analysis also readily revealed the presence of a number of specific compounds in the three brands of condom residues analyzed (i.e. LifeStyles, Trojan and SKYN), including octylamine, palmitic and linoleic acids, urea, *N*-methylmorpholine, oleamide and others.

Despite the progress that has been made in classifying condom residues according to broad categories based on the presence of a handful of specific chemical components commonly found in condoms, it has also been recognized that the application of statistical analysis approaches to condom lubricant identification would eliminate the subjectivity associated with visual comparisons of complex spectra [13]. To this end, Mirabelli et al. [14] analyzed 10 condom types by desorption electrospray ionization (DESI)-MS, and easy ambient sonic spray (EASI)-MS. The data were then processed by supervised pattern recognition statistical analysis approaches (LDA—linear discriminant analysis and SIMCA—soft independent modeling of class analogy) to determine the classification rules for distinguishing between condom types. LDA using the EASI-MS data gave 99% prediction ability, while 94% prediction ability was observed using the DESI-MS data for both LDA and SIMCA. Maric et al. [13] performed multivariate statistical analysis, including hierarchical clustering analysis (HCA), principal component analysis (PCA), and LDA, of DART-MS data generated from 20 condoms (in addition to several other bottled lubricants). High accuracy rates were observed, with discrimination being achieved based on key chemical components and minor additives including lidocaine and benzocaine (anesthetics), spermicide (N-9) and slip agents among other compounds. Spencer et al. [15] used a combination of MALDI-MS, ATR-IR spectroscopy and chemometrics to differentiate condom residues and allow their trace detection in the presence and absence of biological fluids. The biological fluids were found to impose a suppressive matrix effect on the spectra in many cases, although this could be countered to some extent by the use of a cationization agent (NaCl) in the applied matrix. The peaks representing the small molecule additives which give the condoms their distinguishing features (such as numbing and warming agents) were often masked by the matrices used (e.g. dihydroxybenzoic acid and α-cyano-4-hydroxycinnamic acid) unless the cationization agent was used. PCA allowed classification of condom lubricants according to whether they were PEG-based or PDMS-based, with and without small-molecule additives.

The aforementioned reports demonstrate proof-of-principle that condom-derived residues can be identified as such, and that their presence can be detected even when incorporated within biological matrices such as vaginal fluid [1,3,8]. Nevertheless, in order for condom residue identification to find broad utility in a forensics context, a number of issues remain to be addressed. One of them is that the range of types and brands of condoms that have been included in any single study have been relatively small. The largest number of types of condoms that have appeared in a study to date is 58 [8]. In addition, only a narrow range of brands have been represented in any one study. Thus, it has yet to be shown that a large number of types of condoms sourced from around the world can be efficiently differentiated. A second issue is that most of the aforementioned methods involve significant sample preparation steps and utilize amounts of sample that may be deemed too precious to sacrifice. Although it has been recommended that this challenge can be circumvented by collecting duplicate samples during crime scene investigation exercises (for example, collecting two swabs of evidence as opposed to only one, so that the second can be devoted to analysis of non-DNA evidence), this suggestion has not caught on. A third is that several of the approaches that provide the most discrimination between condom types involve the analysis of a composite of data types acquired using multiple instrumental techniques, which is time consuming and presupposes the availability of all of them. Fourth, the majority of methods still only differentiate condoms by broad categories, such as whether they contain PDMS, PEG, N-9 etc., rather than by brand. This limits the usefulness of the findings since compounds such as PDMS, PEG etc. are present in hundreds of types of condom lubricants and other products. Fifth, in order to create a database of condom residues against which unknown condom residues can be screened and identified, it will be necessary to generate a large enough number of sample replicates in order to make multivariate statistical analysis feasible. In principle, a large and continuously increasing dataset comprised of a diversity of samples will provide a more highly representative database. The large number of replicates required is generally difficult to acquire rapidly because of the sample preparation steps, protocols and instrumental methods development required. Sixth, it is essential that markers that enable distinctions to be made between brands be identified. Such findings would permit more accurate brand identification, and provide the opportunity for studies aimed at identifying these compounds within relevant biological matrices, and determinations of their lifetimes.

Towards the goal of addressing some of the aforementioned concerns, we explored the utility of DART-HRMS for the rapid analysis of 110 different types of condoms representing 16 brands from around the globe, in order to generate a database of condom residue spectra that served as representative fingerprints for each brand. No sample preparation steps were required and residues could be analyzed directly with results produced within a few seconds. Application of a variety of statistical analysis steps to the data revealed a sequence that permits the identification of brands with an accuracy of 97.4%, and *m/z* values representative of markers that enable discrimination between brands

were identified. The impact of time since deposition, the presence of contaminants and the influence of transfer of the residue to an alternative surface, on the prediction ability of the model were also assessed. It was found that the method retains the ability to predict sample identity even with aged and dust-contaminated samples, but it is less reliable for identification of residue that has been transferred to another surface, such as glass.

2. Materials and methods

2.1. Instrumentation

An AccuTOF-DART (JEOL USA, Inc, Peabody MA, USA) high resolution mass spectrometer with a resolving power of 6000 FWHM (full width at half maximum) was used for mass measurements. All analyses were performed under soft ionization conditions with the following settings: orifice 1, 20 V; orifice 2 and ring lens voltages, 5 V; grid voltage, + 50 V; peaks voltage, 600 V (in order to detect ions greater than *m/z* 60). The DART simplified voltage and pressure (SVP) ion source (IonSense Inc., Saugus, MA, USA) was operated using a helium gas heater with the temperature set at 350 °C and a flow rate of 2 L min^{-1}. All experiments were conducted in positive ion mode.

2.2. Materials

Condoms were purchased from Amazon (http://www.amazon.com, USA), Walmart (Albany, NY, USA), and JingDong Mall (http://www.jd.com, China). A total of one-hundred and ten different condom types were tested, representing the following sixteen brands: Aoni, Atlas, Caution Wear, Crown, Durex, Fantasy, Glyde, Jissbon, Kimono, LifeStyles, Mates, Now, Okamoto, One, Pasante and Trojan (listed in Table 1). These brands were manufactured in the United States, United Kingdom, Australia, Japan and China.

2.3. Methods

To train the discrimination model, condom-derived lubricants were sampled and analyzed by swiping the closed end of a melting point capillary tube (VWR, Radnor, PA, USA) on the outer surface of a rolled condom and suspending the tube in the space between the DART ion source and the mass spectrometer inlet for between five and ten seconds. Analyses were performed in replicates of either five or ten. All condoms were analyzed in the same manner. Polyethylene glycol (PEG 600) was analyzed with every acquired spectrum as a mass calibrant for accurate mass determinations.

Several experiments were designed to investigate whether the trained model was able to accurately identify the condom brand when the condom residue was exposed to open air for several days, was contaminated with dust, or transferred to a surface such as glass. To determine the impact of aging and exposure to open air, pairs of condoms from the representative brands Atlas and Kimono were used. Within each pair, one condom was unrolled while the other was retained in rolled form as a control. They were then analyzed by DART-HRMS on days 1, 2, 3, 4, 5, 12 and 19 in replicates of five. To determine the impact of external contaminants, the three representative condom brands tested were Caution Wear, Fantasy and One. Two of each of these condoms were analyzed: one in its native, rolled form to act as a control and the second, rolled but contaminated with dust. To contaminate the condom with dust, the index and middle finger of a gloved hand were used to swipe, in a single pass, a 15 × 5 cm area of a window sill that had been allowed to accumulate dust undisturbed for ~6 months. The thumb, index and middle finger of the gloved hand were then used to sprinkle the dust onto the condom sample. This process was repeated to produce each of the dust-contaminated samples.

To determine the impact on their spectra of transferring the lubricants to another surface, one condom from each of the brands Caution Wear, Fantasy and One was wiped across the surfaces of two glass slides to simulate lubricant transfer. One slide acted as a control while the second was contaminated with dust, as previously described. Each of the four samples for each brand was analyzed by DART-HRMS on days 1, 2, 3, 4 and 5 in replicates of five, and one replicate was analyzed on day-12.

2.4. Mass spectral data processing

TSSPro3 software (Shrader Analytical, Detroit, MI, USA) was used for data processing of mass spectra including mass spectral calibration, averaging, background subtraction and peak centroiding. The resulting spectra were exported as text files for statistical analysis processing using the Classification [16] and SpaSM toolboxes [17,18] of the program MATLAB (The MathWorks, Inc.).

2.5. Multivariate statistical analysis of DART-HRMS-derived chemical fingerprints

Multivariate statistical analysis methods were applied to the mass spectral data derived from 110 condom types representing 16 condom brands (listed in Table 1). The sequence of steps (1 through 3) is illustrated in Fig. 1 and described below.

Step 1 (Raw data collection): DART-HRMS spectra representative of each condom sample were measured, processed using TSSPro3 software, saved in text format and imported into MATLAB 9.3.0, R2017b Software (The MathWorks, Inc., Natick, MA, USA) for further analysis.

Step 2 (Data pre-processing): the spectra were binned (i.e. arranged in a matrix with common *m/z* values aligned using in-house written codes in MATLAB). In the binning process, the optimal bin width was selected based on the resolution of the JEOL AccuTOF-high-resolution mass spectrometer used (i.e. ± 5 mmu of the calculated mass), and the threshold abundance was set to 0.1% of the relative peak abundance. Next, uninformative data (i.e. *m/z* values that were present in less than 0.7% of the samples *and* which were of < 5% relative intensity) were removed. This resulted in a 725 × 2104 data matrix, where the 725 refers to the number of spectra (i.e. observations) and 2104 the number of *m/z* values (variables) in the range of *m/z* 59–850. Column-wise normalization was accomplished using the "*autoscaling*" function to provide a standard deviation of 1 for all the variables. As the data matrix contained more variables (2104) than observations (725), a circumstance which could potentially reduce the accuracy of the statistical analysis (because of the presence of significant noise, for example), a minimum redundancy maximal relevance (mRMR) [19] feature selection algorithm was applied to remove irrelevant and dependent variables. This reduced the dimensions of the data to a 725 × 700 matrix.

Step 3 (Classification and marker determination): the data were treated in two ways. The purpose of the first treatment was to develop a model for brand identification. To accomplish this, the matrix was partitioned into a training set and a validation set containing 608 and 117 samples respectively. The number of latent variables (LV) for the PLS-DA [16] was set to 40 components, as this was found to be optimal. The PLS-DA classification model was then tested with the 117 samples in the validation set. A conformal predictor was then applied to determine the class assignment and the precise levels of confidence of new predictions. Confidence values are an indication of how likely it is that a given prediction is correct. Conformal prediction uses past experience to determine a reasonable confidence level to guarantee a maximum error rate for the predictions [20,21]. The details of the conformal predictor algorithm are shown in Scheme 1. In the algorithm, bootstrap sampling is done to provide several (1000×) training and calibration sets so that there are enough samples included in the calibration sample set (i.e. bag) to provide a confidence level. A non-conformity measure α is defined based on the regression properties of the PLS-DA algorithm. Non-conformity measure values indicate how different a given sample

Table 1
Brands and types of condoms studied along with their lot numbers.

Condom Brand	Condom Type	Lot Number	Condom Brand	Condom Type	Lot Number
Aoni	Extra Smooth Ultrathin 001	D15AE21210	**Okamoto**	003 Aloe	HK026B100
	Nanosilver Ultrathin 001	D15AE20704		003 Hyaluronic Acid	HK016A101
	Overtime	D15AE10501		003 Platinum	HK016C98
	Ultrathin	D15AE20703		Charm	HK016A38
Atlas	Blue	16N762		Crown	175J1101
	Extra Large	16N760		Roman	145L1101
	Non-Lubricated	15X4272		Ultra Smooth	HK015M39
	Purple	16N762	**One**	576 Sensation	15N3752
	Red	16N762			15N4269
	Studded	15N2886		Aqua	14N1357A
	True Fit	15N1970			14N1367A
	Ultra-Lubed	16N758			16N950A
	Ultra Thin	15N2151		Banana Split	15F1204
		15SDPN514		Black	14N1357K
	Yellow	16N762			14N1367K
Caution Wear	Black Ice	UT27322			16N950K
	Classic	PN27325		Blue	16N950B
	Mission 707	DN21332		Bubblegum	15F1201
	Wild Rose	RN27322			16F1207
Crown	Skinless Skin	T455		Chocolate Strawberry	15F1208
Durex	Extra Sensitive	1000079540		Classic Select	14N1579
		1000092687		Emerald	16N950E
	Love	1000106912		Fresh Mint	15F1206
	Tropical Apple	15F4190A		Glowing Pleasures	15N858
	Tropical Banana	15F4190B		Green	16N950G
	Tropical Strawberry	15F4190S		Island Punch	15F1197
Fantasy	Banana	16F725			15F1203
	Chocolate	16F726		Lavender	16N950L
	Grape	16F727		Mint Chocolate	15F1205
	Lubricated	5044		Orange	16N950O
	Mint	13F2248		Pleasure Dome	15N1673
		16F728		Pleasure Plus	14N2282
	Strawberry	16F729			14N2284
	Vanilla	16F730		Purple	16N950P
Glyde	Blueberry	BB22551		Red	16N950R
	Cola	BL21131		Super Sensitive	15N846
	Strawberry	PS20521		Tantric Pleasures Maori	14N1403
	Vanilla	PV23801		Tantric Pleasures Titan	14N1402
	Wildberry	PW23801		Tantric Pleasures Tribal	14N1401
Jissbon	Super Moist	JP150610		The Legend	15N1215
	Ultra Thin	JB150618		Vanish	15N474
	Zero	1507ZCRL91			15N475
Kimono	Micro Thin	50166-9		Yellow	16N950Y
	Micro Thin Large	50168-9		Zero Thin	1171501
	Micro Thin plus Aqua Lube	50451-9			1171502
	Ribbed + Sensi Dots	40554-9	**Pasante**	Blueberry Blast	PL5437B
	Thin	50461-9		Chocolate Temptation	PL5437C
LifeStyles	Red	1505991922		Mint Tingle	PL5437G
	SKYN Extra Lube	1505P10622		Strawberry Crush	PL5437R
	SKYN Extra Studded	1511843316	**Trojan**	ENZ	0T4277X1
	SKYN Original	1507P10722		Her Pleasure Sensations	TT5153UZ922
	Ultra Sensitive	1507130416		Intense	TT5146UZ811
	Ultra Thin	1505020422		Magnum Ecstasy	TT4302CB
	Yellow	1505991922		Magnum Lubricated	TT5054XZ523
Mates	Banana	1412341216		Magnum Thin	TT5251ZZ1216
	Mint	1410421216		Magnum Warming	TT4121ZZ516
	Strawberry	1411751216		Twisted	TT5104WZ718
	Vanilla	1411741316		Ultra Ribbed	TT5301Y
Now	Carnival Banana	1404004-4		Ultra Thin	TT5305BZ110
	Carnival Mint	1404004-3		Warming	TT5157UZ1302
	Roller Coaster	1404002			
	Speed Bumps	1404003			
	Super Fine	1404001			

is from the samples in a bag.

In each iteration of the algorithm and after every prediction that is made for each new instance of the calibration set, the instance is re-trained against the underlying model. The algorithm provides prediction validity values (p-values) for each test condom. If the p-value for one/several brand(s) is > than the assumed error, the test condom is accepted as a member of the brand(s). The outputs of this treatment are confidence level (i.e. 1 minus the second largest p-value), and credibility (i.e. the largest p-value) respectively, if single label prediction is desirable.

The purpose of the second treatment was to identify the markers that were the most important contributors to the ability to discriminate between brands. To accomplish this, the original data were reanalyzed with bootstrapping (1000×), which involved the random sampling of 90% of all observations each time (i.e. 649 variables each time). Within each bootstrapping iteration, the re-sampled data were subjected to

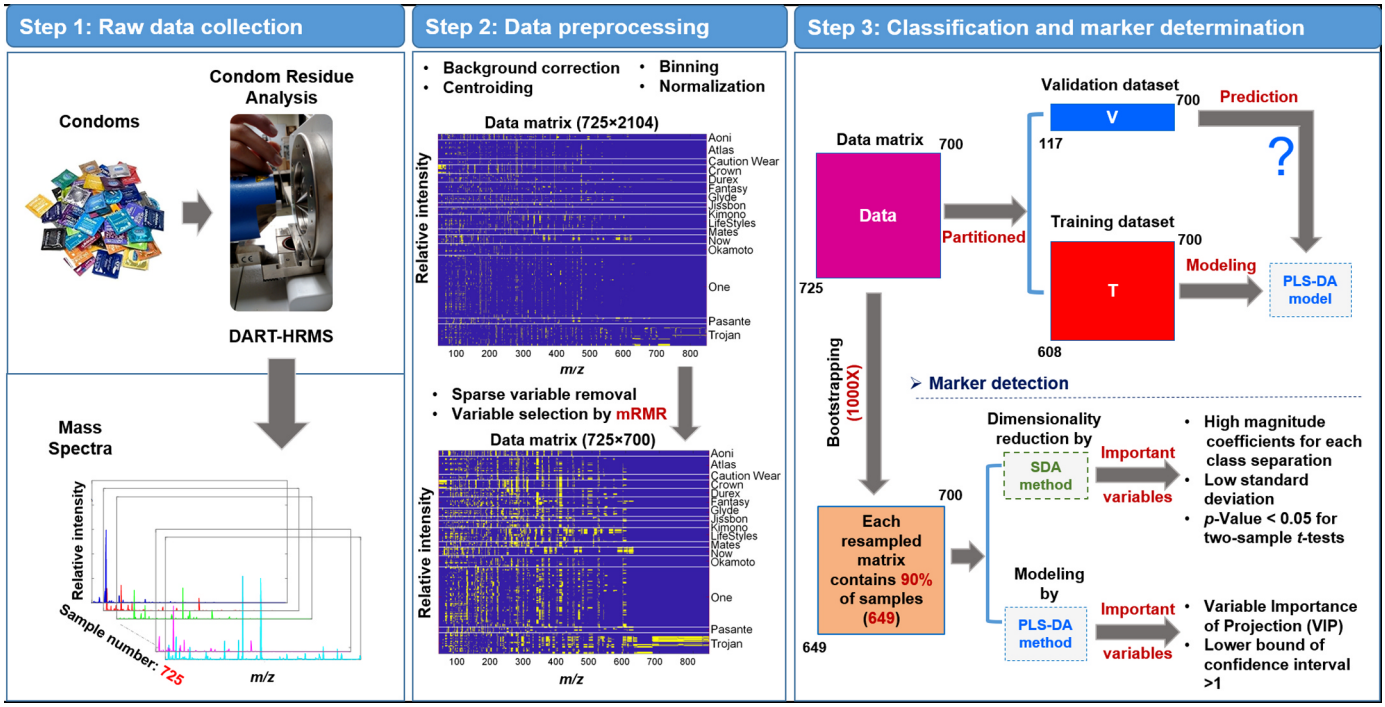


Fig. 1. Steps in the multivariate statistical analyses of DART-HRMS-derived chemical fingerprints of condom lubricant residues.

either PLS-DA or SDA to compute the variable importance in projection (VIP) [22] scores and the Sparse Discriminant Analysis (SDA) [17,18] coefficients. On completion of the bootstrap iterations, the averages and standard deviations of the generated VIP scores and SDA coefficients were determined.

2.5.1. PLS-DA

VIP scores from this analysis were used to detect the most effective variables for discrimination. They were identified using the 95% lower-bound bootstrapped confidence interval for the mean of VIP scores of > 1. The 95% confidence intervals for the mean scores were calculated by multiplying the standard deviation of each variable's scores by the appropriate quantile from the *t*-distribution (in this case 1.6448).

2.5.2. SDA

The dimensionality of the data was reduced through SDA. In this analysis, the averages of the coefficients (from 1000 repetitions of the bootstrapping method) were used to calculate the discriminative scores. Then, the specific direction that was representative of each class (i.e. brand) differentiation was determined and used to reveal the most significant variables for discrimination of that class from the others. The corresponding coefficient vector and SDA score was fit to a fixed range to find the *m/z* values that had coefficients that correlated with the separated class scores.

3. Results and discussion

3.1. Differentiation within brands

The diversity of condoms analyzed in this study is listed in Table 1. A total of 110 condom types representing 16 brands were analyzed. Within each brand, condoms could be classified according to various features including the physical characteristics of the rubber itself, or the presence of: (1) specialty chemicals designed to elicit a specific sensory response such as numbing agents; (2) flavor chemicals such as those associated with the taste of banana, chocolate, grape, etc.; and (3) a spermicidal agent such as N-9. The condom lubricants were easily and rapidly sampled by DART-HRMS in positive ion mode under soft ionization conditions. No sample pretreatment was involved and each analysis was accomplished in less than 5 s. The analyses furnished spectra with peaks representing the protonated un-fragmented molecules that were detected. A subset of representative spectra illustrating one condom from each brand, and featuring special characteristics of the rubber (i.e. Atlas Studded and LifeStyles Ultra Thin), flavoring agents (i.e. Mates Vanilla and Fantasy Banana) and other attributes, is shown in Fig. 2, while the full complement of spectra acquired is shown in Supplementary material Fig. S1. Compounds with masses consistent with those that would be expected to be present in condoms advertised to have flavor characteristics were observed in all cases and are shown for some of these in Fig. 2. For example, the Mates Vanilla spectrum had a mass at *m/z* 153.0552 which is consistent with the presence of protonated vanillin, the molecule primarily responsible for the flavor and odor of vanilla. Fantasy Banana had a mass at *m/z* 131.1072, consistent with the presence of protonated isoamylacetate, a primary flavor and odor constituent of bananas. Pasante Mint Tingle had a mass at *m/z* 151.1113 which is consistent with the presence of protonated carvone, a primary constituent of mint flavoring. Other tentatively identified molecules that are commonly found in condoms are indicated in Fig. 2.

3.2. Discrimination between brands

An important attribute of the types of data that can be classified and thus differentiated using statistical analysis methods is that there is consistency among like samples. To assess this, the DART-HRMS spectra of several samples of the same type of condom but with different lot numbers were compared in head-to-tail plots of their mass spectra. Supplementary material Fig. S2 shows representative examples featuring mass spectra for lot number comparisons of Atlas Ultra Thin, Durex Extra Sensitive, Fantasy Mint and One 576 Sensations. The results show that in every case, the spectra are very similar to one another, indicating that the chemical fingerprints remain consistent between batches.

Numerous types of condoms under various brands are advertised to have the same characteristics. This is particularly true of flavored varieties. For example, brands Fantasy, Mates, Now, One and Pasante all have mint-flavored condoms. Therefore, the question of whether

Input: For a data set: $\{(x_1, l_1), \ldots, (x_{ns}, l_{ns})\}$, define matrices $X(ns \times nv)$ and $Y(ns \times nc)$, where ns, nv, l and nc represent the number of samples, variables, class assignments and number of classes respectively. Each column of Y contains binary values that indicate the presence or absence of a sample in a class by 1 and 0 respectively.

Perform bootstrap sampling and provide q number of training and calibration sets, designated as t and c respectively.

The bootstrap output is: q training sets: $\{(x_1, l_1), \ldots, (x_{ns-m}, l_{ns-m})\}_t$ and q calibration sets: $\{(x_1, l_1), \ldots, (x_m, l_m)\}_c$,

For i=1 to q

Train the PLS-DA model using: $\{(x_1, l_1), \ldots, (x_{ns-m}, l_{ns-m})\}_{ti}$

Predict the first sample, $\{(x_1, l_1)\}_{ci}$, of the calibration set

For j=1 to m

Calculate the non-conformity measure, $\alpha_{ij}^k = \frac{|1-\hat{y}_{ij(k)}|}{\sigma_i}$, where $k \in \{1, \ldots, nc\}$, σ_i defines the accuracy of the model i, and $\hat{y}_{ij}$ is the result of the PLS-DA prediction

Train the PLS-DA model using: $\{(x_1, l_1), \ldots, (x_{ns-m}, l_{ns-m})\}_{ti}$ (training samples), and $\{(x_{1:j}, l_{1:j})\}_{ci}$(calibration samples)

Predict the sample using: $\{(x_{j+1}, l_{j+1})\}_{ci}$

end

end

For each test sample x_{nb+1}, where $nb = q \times m$, perform the prediction and calculate the non-conformity measure, α_{nb+1}^k

Calculate the prediction validity using: $p(\alpha_{nb+1}^k) = \frac{count\{r \in \{1,\ldots,nb+1\} | l_r = k \,\&\, \alpha_{nb+1}^k \leq \alpha_r^k\}}{count\{r \in \{1,\ldots,nb+1\} | l_r = k\}}$

Output: If a single-label prediction is desirable, the confidence of this prediction is one minus the second largest p-value, and the credibility is the largest p-value.

Scheme 1. Algorithm applied for the computation of the conformal prediction used to determine the confidence levels of new predictions.

condoms of different brands but with the same advertised characteristics could be distinguished was investigated. Fig. 3 shows a collection of the spectra of six types of condoms representing the five brands (i.e. Fantasy, Mates, Now, One and Pasante) that are all advertised to be mint flavored. In each panel, the full mass spectrum is shown in the inset, while the area of the spectrum where the mint flavoring agents carvone and thujone would appear (*m/z* 151.1123 and 153.1280 respectively when protonated) is magnified for clarity. Carvone and thujone were identified in the mass spectra of all samples except Now Carnival Mint. However, although there were peaks that were common to most of the spectra, the overall chemical fingerprints of the spectra were nevertheless unique. This indicates that even brands marketed as having the same lubricant characteristics have spectra that can be visually distinguished.

In order to assess whether the mass spectral data could be used to classify and by extension identify the condom residues, the > 700 spectra representative of 110 condom types under 16 brands were subjected to statistical analysis processing. A summary of the steps that were applied is outlined in Fig. 1. Among them were Sparse Discriminant Analysis (SDA) [17,18] which was implemented to visualize the similarities and differences between the condom samples, and PLS-DA [16], which was used to create a model for discriminating between the brands.

Exploratory analysis of the similarities among the condom data (rendered as a matrix) was performed by SDA. This method accomplished simultaneous classification and feature selection with a sparseness criterion imposed, and reduced the data to 15 vectors (termed SD_1 through SD_{15}) which contained the information most responsible for discrimination. Each of these SDA discriminant vectors was defined by < 300 non-zero features. Next, SDA scores plots featuring the various total numbers of binary combinations of the 15 vectors (i.e. 210 in total) were created to assess which subset enabled separation of the various condom classes. Fig. 4 shows the subset of projections of the condom data onto the SDA most discriminant vectors that revealed class assignments (with each brand represented by a different color). The separated brands in each subspace are indicated by name.

To develop a system that could be used to predict the brand identity of condom residues, the discrimination model was trained using the PLS-DA algorithm on the 725 × 700 matrix as described in the Materials and methods section. The number of latent variables (LV) for PLS-DA was set to 40 components, as this was found to be optimal. The cross validation (*100-fold* venetian blind) and bootstrapping (with 1000 repetitions) techniques were applied to validate the model and overcome the risk of overfitting. The performance of the model is reported as a confusion matrix (Fig. 5). It shows that the 40 LV explained 65% of the variance of the predictor (X) and 80% of the response variance (Y), and created a classifier with an accuracy of 97.4% in cross validation.

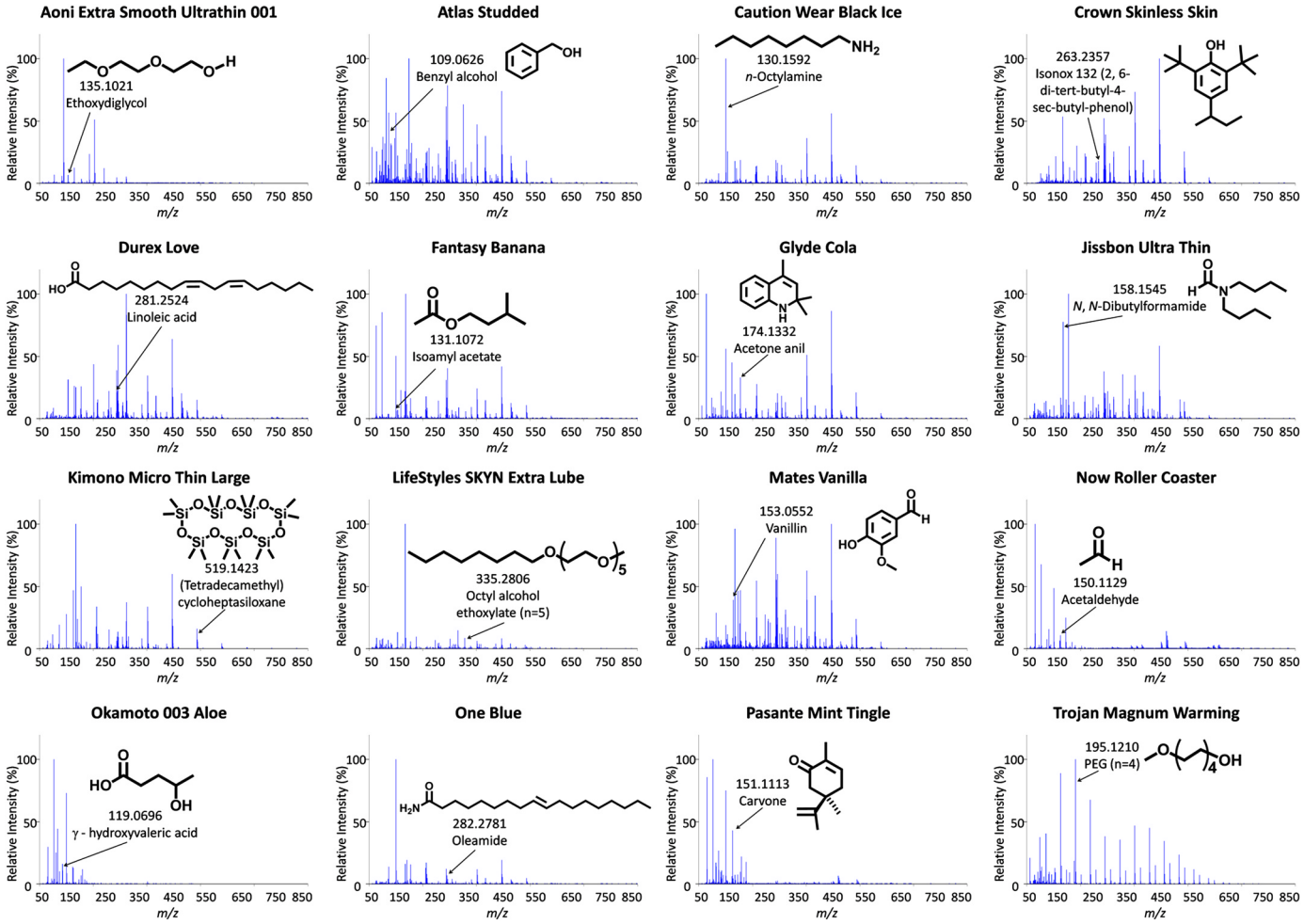


Fig. 2. A subset of representative examples of DART-high resolution mass spectra of condom lubricants representative of the 16 brands investigated in this study. One spectrum for each of the brands studied is presented, and each panel indicates the type of condom represented in the spectrum. The spectra were acquired under soft ionization conditions in positive ion mode. Thus, the peaks represent the protonated forms of the detected molecules. The structures of tentatively identified molecules are shown.

The diagonal elements of the confusion matrix show the number of correct predictions for each class and all off diagonal elements represent misclassified data. For each class, the percent true and false classification rates are shown in the un-shaded areas and are indicated in red and brown colors respectively. The classification precision, sensitivity and specificity are also illustrated for each class. The confusion matrix revealed that there were 9 brands which were predicted with 100% accuracy - Aoni, Crown, Durex, Glyde, Kimono, LifeStyles, Now, Okamoto and Pasante. The performance accuracies for Atlas, Caution Wear, Fantasy, Jissbon, Mates, One and Trojan were 95.1%, 85.0%, 94.6%, 92.9%, 95.0%, 97.1% and 98.3% respectively. The PLS-DA classification model was then tested with the 117 samples in the validation set and the conformal predictor algorithm (Scheme 1) was applied to predict the condom brand. The prediction results using a prediction confidence level of 85% (i.e. the assumed error was 0.15 for acceptance of the class label), are presented in the "Prediction" column of Table 2. Of the 117 samples, the PLS-DA misclassified the brand class of 3 of them (indicated in red): 1 sample in the brand Now that was classified as Kimono; 1 sample in the brand LifeStyles that was classified as Mates; and 1 sample in the brand One that was identified as Caution Wear. For each of the randomly selected samples that appear in the table, the *p*-values for the credibility, and the confidence levels of the predictions of the class with highest *p*-value (i.e. most probable class) are shown.

3.3. Marker detection

Two approaches were used for marker detection. The first was PLS-DA modeling, and the second, the SDA method. For PLS-DA, the variable importance in projection (VIP) [22] scores obtained by bootstrapped PLS-DA were used to reveal 68 variables (*m/z* values) that were the most important for discrimination. These were identified using the 95% lower bounds of the bootstrap confidence interval for the mean VIP scores of > 1, and they appear above the green line in Fig. 6, spanning a mass range from nominal *m/z* 71–519. The horizontal green line in Fig. 6 represents the VIP scores threshold value (i.e. 1). The figure reveals that *m/z* 130 had the highest relative VIP score, followed by *m/z* 445, 93 and 158 (the latter two of which were of similar magnitude). These high VIP scores indicate that these masses were major contributors to the total ion counts in the mass spectra in which they were observed. Visual inspection of the spectra (Figs. S3 and S4) showed them to not only be major peaks, but also to be present in most of the brands tested. However, *m/z* 130 was absent in Crown (1 condom brand represented by 25 spectra), *m/z* 158 was absent in Aoni and Crown, and *m/z* 445 was absent in Aoni and Now. In addition, the total ion counts for *m/z* 130 across the Aoni, LifeStyles and Trojan brands were low. The counts for *m/z* 445 were very low for Pasante. These characteristics are illustrated in Fig. S4, which displays the relative total ion counts across all brands for nominal *m/z* 122, 130, 158, 174, 227, 263 and 445. The figure also shows the unique markers *m/z* 122 and

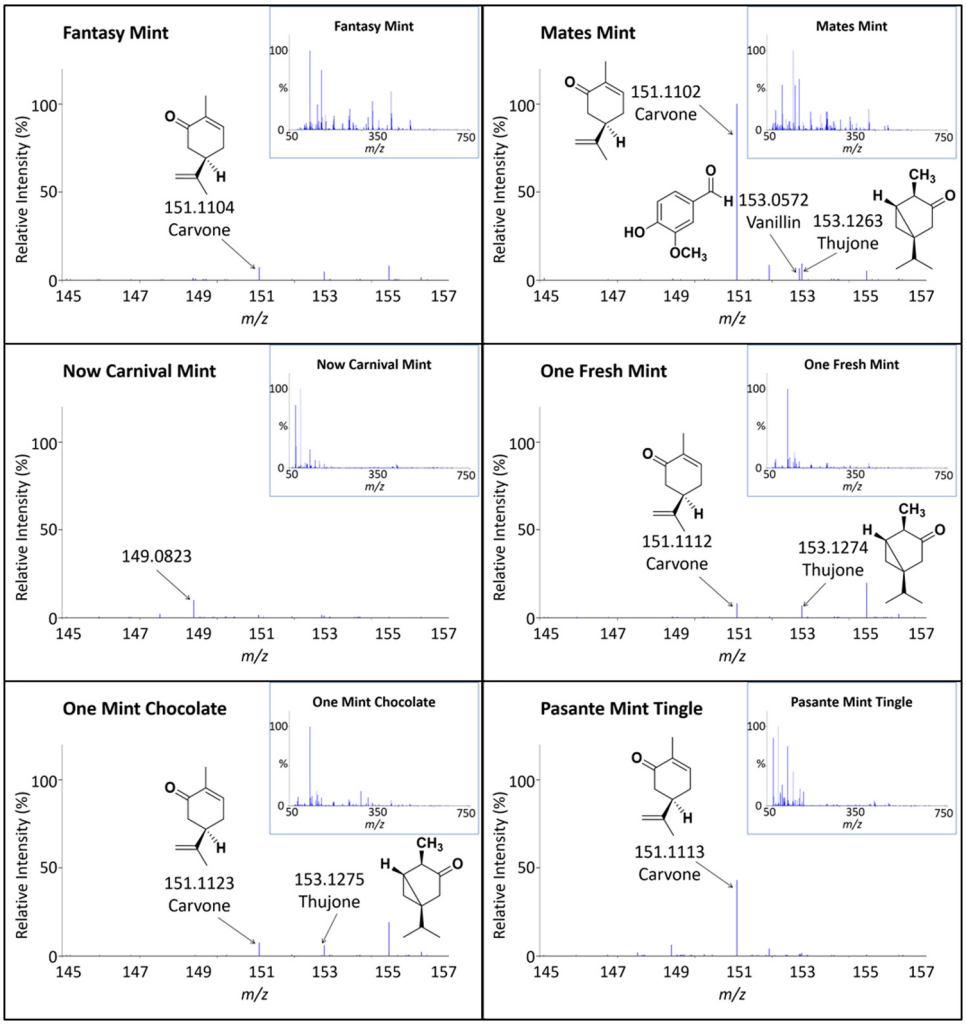


Fig. 3. Comparison of the spectra of different brands of condoms all advertised to be mint flavored. In each panel, the area where the high-resolution masses of mint flavor and odor chemicals (i.e. carvone and thujone) appear magnified, while the full spectrum is shown in the inset. A peak tentatively identified as vanillin, is highlighted in the Mates Mint spectrum. Notably, carvone and thujone were not detected in the Now Carnival Mint samples.

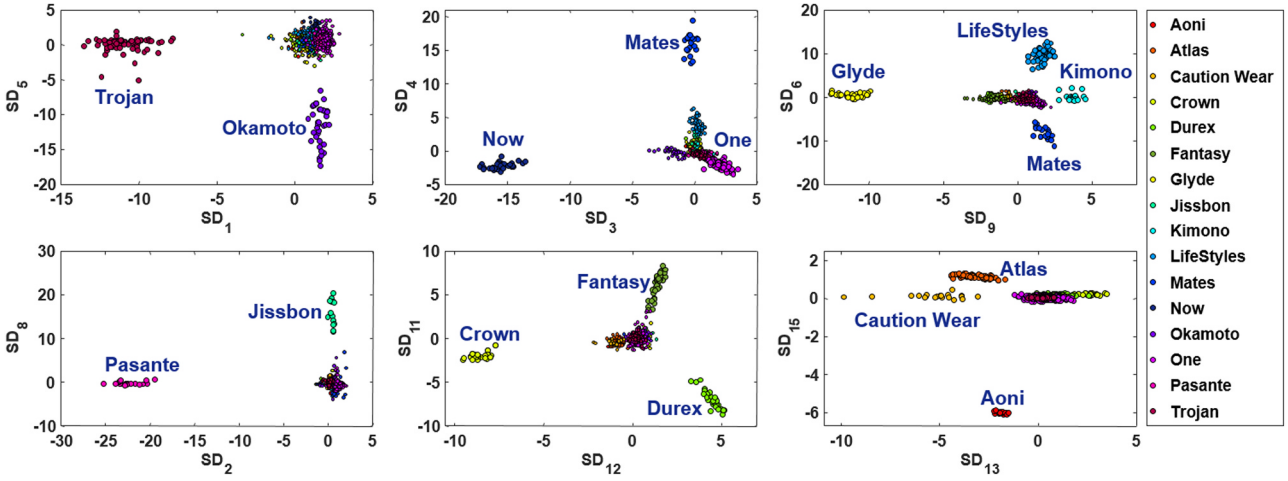
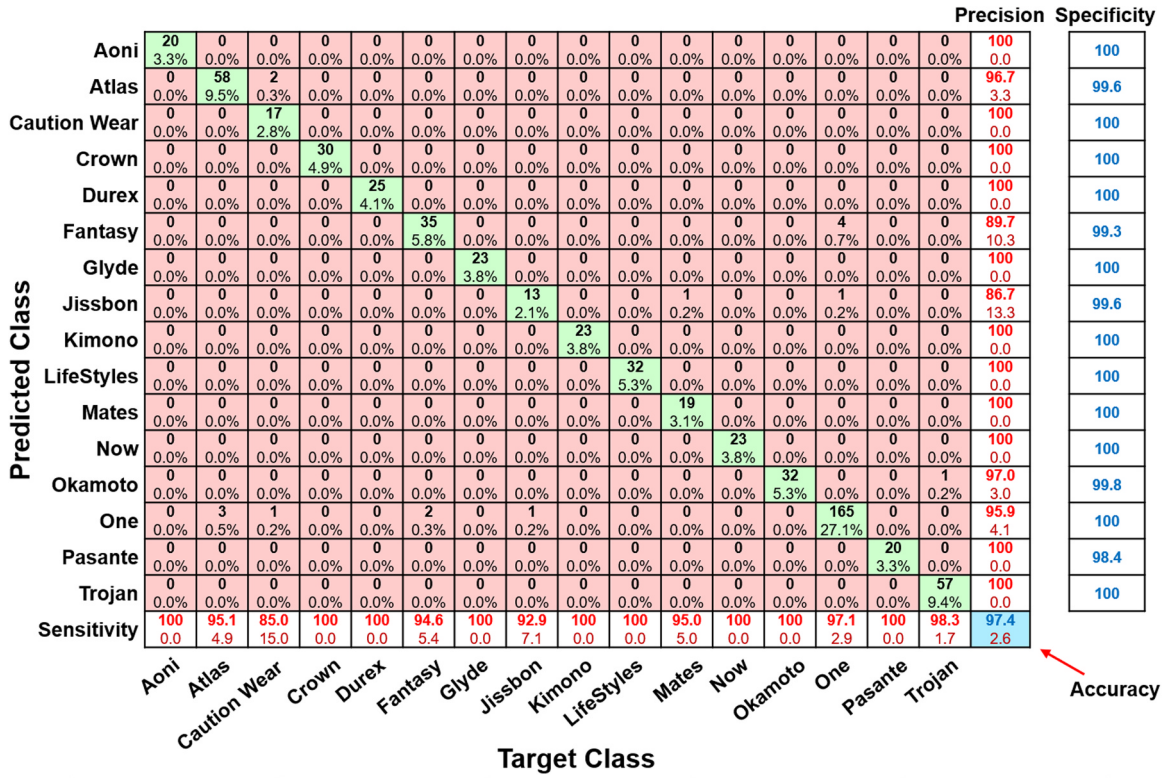


Fig. 4. Projection of the condom data onto the SDA most discriminant vectors to reveal class assignments (with each brand represented by a different color). Thus, the Trojan and Okamoto classes emerged separated from other classes in the plot of SD_1 against SD_5; Mates, Now and One could be distinguished from other brands in the plot of SD_3 versus SD_4; LifeStyles, Glyde, Kimono and Mates were revealed in the plot of SD_9 against SD_6; Jissbon and Pasante could be distinguished in the plot of SD_2 versus SD_8; Crown, Durex and Fantasy could be differentiated in the plot of SD_{12} against SD_{11}; and the plot of SD_{13} versus SD_{15} revealed separation of Caution Wear, Aoni and Atlas brands from the others.



Number of LV	Explained variance of predictor (X)	Explained variance of response (Y)	Model Error	Cross Validation Error
40	65%	80%	0.0035	0.0265

Fig. 5. Results of the validation of the PLS-DA model in discrimination of condom brands using the *100-fold* venetian blind cross validation method as revealed by a confusion matrix. The diagonal elements show how many of the observations are correctly estimated by the model. The off diagonal elements show misclassifications. The column on the right of the plot lists the accuracy for each predicted class, while the row at the bottom of the plot shows the accuracy for each true class. The cell shaded turquoise at the bottom right of the plot shows the overall accuracy. The PLS-DA model specifications, i.e. the measure of the receiver operating characteristics for each output class [precision, sensitivity (true positive rate) and specificity (1 minus the false positive rate)] are illustrated as well. The table beneath the confusion matrix displays the number of optimal latent variables used for the model, the variance of the response and the predictors that defined the 40 latent variables, and the model and cross validation errors.

227 for Aoni and Pasante respectively.

One question that arose in the course of this study is whether "universal" condom residue markers exist. Such compounds would be those whose presence provides strong evidence or "proof" that a trace material is condom-lubricant derived. To be useful, they would need to be present in most, if not all condoms. If such molecules exist, they could be used to develop an on-site analysis approach that could be used to determine whether trace evidence found by crime scene investigators is condom-derived. In this regard, a subset of 14 of the *m/z* values represented in Fig. 6 were observed in all condom brands. These were nominal *m/z* 73, 83, 89, 91, 97, 117, 135, 256, 280, 282, 284, 372, 373 and 397 and in principle, any of these could serve as a universal condom residue marker. Of these, *m/z* 89, 91 and 117 corresponded to the protonated forms of formulas $C_4H_8O_2$, $C_4H_{10}S$ and $C_6H_{12}O_2$ respectively. Depending upon the condom, formula $C_4H_8O_2$ could be acetoin, ethyl acetate, or methyl propionate, all of which are known flavor and olfactory agents [23,24] added to condom lubricants. Similarly, formula $C_4H_{10}S$ could be 2-methyl-2-propane thiol or butane-1-thiol, both of which are reported to be additives in some condom lubricants [24]. Therefore, these *m/z* values would not serve as good condom residue markers, because they will have different molecular identities depending upon the brand. In addition, these compounds are ubiquitous and commonly found in food and other products. Oleamide (corresponding to *m/z* 282 in its protonated form) is used in condoms as a slip agent, but it is also found as a component of polypropylene used in plastics. In addition, it has been shown to be an endogenous molecule found in trace amounts in cerebrospinal fluid and human blood plasma [25]. Nevertheless, it may prove to be a good marker for condom residues because its levels in condoms are significantly greater than those observed in other matrices where it may exist. Also detected were a subset of compounds (in their protonated forms) that were present in the vast majority of brands (Fig. S4). These included *N,N*-dibutylformamide, with an *m/z* value of 158, which was observed in all but the Aoni, Crown and Now brands; isonox 132, with an *m/z* value of 263, which was observed in all but the Now and Pasante brands; and acetone anil, with an *m/z* value of 174, which was observed in all brands except Aoni, Crown and Now. These three compounds could therefore also potentially serve as condom markers, with the caveat that although their absence does not necessarily mean that a given sample is not derived from a condom, their presence (particularly in combination) is highly suggestive that a given residue is condom-derived. There were also a subset of *m/z* values that were unique to a given brand. These are listed in Table 3. The brand One was the only case where all the *m/z* values observed in its lubricant spectra were also detected in other brands. The identities of the other unique *m/z* values are not yet known and are the subject of further investigations. Tables S1a–g present additional information about the *p*-values for the masses that were identified as being important for discrimination, and which were computed from the comparison of the mean of a brand/class (in each row), versus the mean of all other brands (as the other class).

For the second approach to marker detection, the dimensionality of the data was reduced through SDA. In this analysis, the average of the

Table 2
Prediction results for the external validation samples (117 in total). The conformal predictor algorithm (Scheme 2) was used to calculate the prediction validity (*p*-values) for each brand. The predictions are based on a confidence level of 85%.

Class*	Prediction§	Credibility‡	2nd Highest *p*-value¶	Class*	Prediction§	Credibility‡	2nd Highest *p*-value¶	Class*	Prediction§	Credibility‡	2nd Highest *p*-value¶
Atlas	Atlas	0.30	0.00	LifeStyles	—	0.02	0.01	One	One	0.86	0.01
Atlas	Atlas	0.89	0.00	Now	Kimono	0.53	0.05	One	One	0.40	0.01
Atlas	Atlas	0.34	0.1	Now	Now	0.51	0.01	One	One	0.54	0.01
Atlas	—	0.13	0.1	Now	Now	0.28	0.15	One	One	0.44	0.01
Durex	Durex	0.30	0.01	Now	Now	0.40	0.01	One	One	0.91	0.01
Durex	Durex	0.65	0.01	Now	Now	0.66	0.01	One	One	0.48	0.01
Durex	Durex	0.20	0.01	Now	Now	0.19	0.01	One	One	0.82	0.01
Durex	—	0.05	0.01	Now	—	0.00	0.03	One	One	0.28	0.01
Durex	—	0.05	0.01	Okamoto	Okamoto	0.38	0.01	One	One	0.46	0.01
Fantasy	—	0.05	0.03	Okamoto	Okamoto	0.98	0.01	One	One	0.77	0.01
Fantasy	Fantasy	0.6	0.01	Okamoto	Okamoto	0.83	0.02	One	One	0.82	0.01
Fantasy	—	0.04	0.01	Okamoto	Okamoto	0.38	0.05	One	Caution Wear	0.27	0.09
Glyde	Glyde	0.23	0.10	Okamoto	—	0.04	0.02	One	One	0.17	0.01
Glyde	Glyde	0.25	0.30	Okamoto	Okamoto	0.77	0.01	One	One	0.34	0.01
Glyde	Glyde	0.23	0.01	Okamoto	Okamoto	0.48	0.01	One	One	0.90	0.01
Glyde	Glyde	0.23	0.01	Okamoto	Okamoto	0.95	0.01	One	One	0.99	0.01
Glyde	Glyde	0.19	0.01	One	One	0.91	0.01	One	One	0.75	0.01
Glyde	Glyde	0.82	0.01	One	—	0.18	0.01	One	One	0.86	0.01
Glyde	Glyde	0.90	0.03	One	One	0.38	0.01	One	One	0.90	0.01
Jissbon	Jissbon	0.49	0.01	One	One	0.30	0.01	One	One	0.23	0.1
Kimono	Kimono	0.71	0.01	One	One	0.43	0.05	One	—	0.00	0.05
Kimono	Kimono	0.82	0.01	One	One	0.26	0.01	One	One	0.30	0.02
LifeStyles	LifeStyles	0.38	0.01	One	One	0.44	0.01	Trojan	—	0.05	0.01
LifeStyles	LifeStyles	0.16	0.15	One	One	0.93	0.01	Trojan	Trojan	0.38	0.01
LifeStyles	LifeStyles	0.58	0.10	One	One	0.38	0.08	Trojan	—	0.16	0.01
LifeStyles	LifeStyles	0.64	0.01	One	One	0.90	0.01	Trojan	—	0.02	0.02
LifeStyles	LifeStyles	0.31	0.14	One	One	1.00	0.01	Trojan	Trojan	0.24	0.01
LifeStyles	LifeStyles	0.76	0.01	One	One	0.27	0.01	Trojan	Trojan	0.33	0.02
LifeStyles	Mates	0.32	0.22	One	One	0.47	0.01	Trojan	Trojan	0.55	0.01
LifeStyles	LifeStyles	0.36	0.03	One	One	0.81	0.01	Trojan	Trojan	0.23	0.01
LifeStyles	LifeStyles	0.64	0.03	One	One	0.85	0.01	Trojan	Trojan	0.39	0.01
LifeStyles	LifeStyles	0.66	0.00	One	One	0.82	0.04	Trojan	Trojan	0.99	0.01
LifeStyles	LifeStyles	0.38	0.00	One	One	0.25	0.01	Trojan	Trojan	0.69	0.01
LifeStyles	LifeStyles	0.42	0.01	One	One	0.16	0.01	Trojan	Trojan	0.89	0.01
LifeStyles	—	0.05	0.07	One	One	0.79	0.01	Trojan	—	0.04	0.01
LifeStyles	LifeStyles	0.31	0.01	One	One	0.70	0.01	Trojan	—	0.05	0.03
LifeStyles	LifeStyles	0.29	0.01	One	One	0.98	0.01	Trojan	Trojan	0.69	0.01
LifeStyles	LifeStyles	0.31	0.01	One	One	0.33	0.13	Trojan	Trojan	0.21	0.01
LifeStyles	LifeStyles	0.14	0.06	One	One	0.34	0.01	Trojan	Trojan	0.73	0.01

*The 117 samples were randomly selected resulting in representation of several condoms within the same brand.
§The predictions are based on an error of 0.15 (i.e. 85% confidence level). The "—" designation indicates that the sample was not assigned to a class (i.e. no class was identified in which the highest *p*-value was > than the error). Entries displayed in red indicate misclassifications.
‡The credibility for the single label prediction is based on the largest *p*-value from among the 16 brands.
¶The confidence of the single label prediction is equivalent to 1 - the second highest *p*-value.

coefficients (from 1000 repetitions of the bootstrapping method) were used to create the discriminative scores. The scores and variable weights were then used to reveal the most significant variables for discriminating each class from the others. Supplementary material Figs. S5–S17 display the sorted averages of the coefficient vectors in each specific direction, along with error bars which illustrate the stability of the impact of each variable in the corresponding direction. The scores are shown in a separate embedded axis (inset) to illustrate the relationship between the discriminated class scores and the coefficients. Classes 1 through 16 which are represented in Figs. S5–S17 refer to brands Aoni, Atlas, Caution Wear, Crown, Durex, Fantasy, Glyde, Jissbon, Kimono, LifeStyles, Mates, Now, Okamoto, One, Pasante and Trojan respectively. Associated with each of the figures is a table that shows information on the coefficients of the 30 variables (listed in the *m/z* columns of each table) which were matched with a direction (positive or negative) of discrimination for the brand in the SDA scores plot (embedded axis—see inset) represented in the figures. Also reported are: the bootstrap estimates of the variables coefficients and standard deviations of the coefficients (with the data in the tables arranged in order of decreasing values of the absolute magnitudes of the

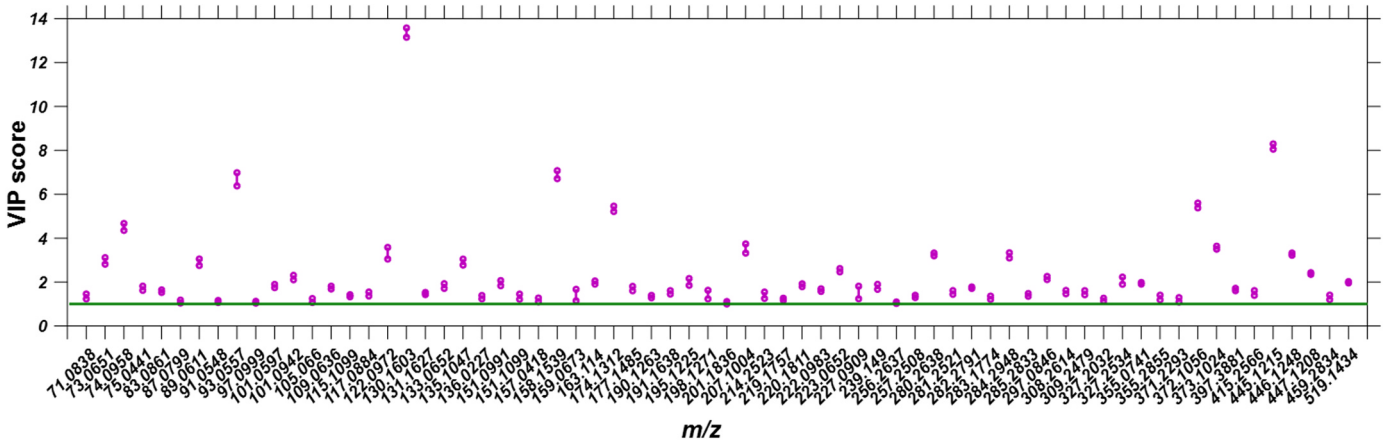


Fig. 6. Variable importance in projection (VIP) scores from PLS-DA, and extracted by the 1000 times bootstrap resampling. They were used to reveal 68 variables (*m/z* values) that were the most important for discrimination. These were identified using the 95% lower bounds of the bootstrap confidence interval for the mean VIP scores of > 1, and they appear above the green line, spanning a mass range from nominal *m/z* 71–519.

Table 3
Subset of unique *m/z* values (± 5 mmu) that serve as potential condom residue markers.

Brand class	Identified *m/z* values	Brand class	Identified *m/z* values
Aoni	122.0972, 123.099, 144.0669, 242.2847	Kimono	187.0968, 229.2181, 403.3634
Atlas	169.486	LifeStyles	173.1355, 183.0801, 259.2409, 279.1027, 377.3309
Caution Wear	285.2981, 312.3250	Mates	129.0924, 147.1016
Crown	212.1155	Now	61.0416, 194.1247
Durex	159.0673	Okamoto	86.0599, 279.2419
Fantasy	152.0719, 165.0926, 348.3312, 349.3355, 489.4061, 516.4319, 544.4652	One	–
Glyde	74.0266, 149.0954	Pasante	123.0912, 127.1059, 128.0442, 128.1088, 142.1230, 227.0909, 264.165
Jissbon	269.2299	Trojan	109.1281, 295.1056

coefficients); and the *p*-values obtained from comparison of the mean of the total ion counts (for the indicated *m/z*) for the given brand, versus the mean (for the indicated ion) for the total ion counts for all the brands combined (using the student's *t*-test). The masses with *p*-values of < 0.05 are highlighted in green, indicating that there is a significant difference between that brand and all the other brands vis-à-vis that variable. The *m/z* values with a coefficient of high magnitude and a low standard deviation, in addition to a *p*-value of < 0.05, represent those variables with the highest impact on condom brand discrimination.

For ease of interpretability, the variance scaled coefficients associated with the variables highlighted in green (within the tables in Supplementary material Figs. S5–S17) are shown with the SDA scores as bi-plots in Supplementary material Figs. S18a–p. The bi-plot is a dual representation scatter plot showing the sample scores as points, and the coefficients with the vectors. It enables visual assessment of a given marker's effect and significance and illustrates which variables are consistent with class separation. Supplementary material Fig. S18a–p provides a more detailed presentation of the SDA scores plots shown in Fig. 4. The *m/z* values in red text represent those highlighted in green in Supplementary material Figs. S5–S17 that are correlated with the separated class and were revealed from scaling of the SDA scores and their corresponding coefficient vectors to a fixed range. The intersection of the red dashed lines in each of these plots represents the position of origin in the 2D Cartesian plot, and the *m/z* values within the dimensional space which are clustered around a class are those revealed to be most heavily weighted in separating the indicated brand(s) from the others. For example, Supplementary material Fig. S18a shows the plot of vector SD_{13} against vector SD_{15} and illustrates that the variables appearing in red, including *m/z* 94, 122, 123, 144, 153 and 242 are most important in distinguishing Aoni from other brands, with 122 being both unique and the base peak in the mass spectrum of Aoni. For the Atlas brand represented in Supplementary material Fig. S18b, important masses for differentiation include *m/z* 130, 169, 190, 372 and 391 among others, with 169.0486 being unique to Atlas. Other masses that were unique to specific brands are listed in Table 3, which illustrates that while some brands exhibited several unique masses, others displayed only 1. Comparison of the results from the PLS-DA and SDA methods indicated that they were consistent in terms of the detected *m/z* values, and they both revealed similar *m/z* values as being important for discrimination (listed in Table S2).

3.4. Effect of external factors on identification accuracy

With the demonstration that chemometric processing of DART-HRMS-derived chemical signatures of condom residues can be used for rapid high-accuracy prediction of condom brands, the question of whether this method could be employed to identify condom residues that had been exposed to open air, or were contaminated, or were transferred to an alternative surface, was considered. To investigate this, condom samples from selected brands that were either exposed to open air for up to 19 days or contaminated with dust and left exposed to open air for up to 12 days, were analyzed by DART-HRMS. The results of the prediction analysis of the spectra using the developed model are displayed in Supplementary material Tables S3–S4 (for Atlas and Kimono brands respectively); Tables S5a/b (for Caution Wear); Tables S6a/b (for Fantasy); and Tables S7a/b (for One). In all of the tables, the credibility values are listed for the brand with the highest *p*-value. The predictions are derived from the confidence level and are based on an 85% cutoff. A single brand prediction is shown in cases where the computed *p*-value for the other brands was lower than the assumed error. Correctly predicted brands are indicated with pink shading and single brand misclassifications are indicated with yellow shading. Multiple brand predictions are shown when more than one brand had a *p*-value greater than the error. The brand with the highest *p*-value is

bolded. The instances where the correct brand was included among the multiple brands that were predicted appear in blue shading, while the instances where multiple brands were predicted but none of them were correct appear in green shading. For the Atlas air exposure experiments (Table S3), the results showed that in the majority of cases, the condom was correctly identified as Atlas. When misidentifications occurred, the predication was usually Caution Wear. For the Kimono air exposure experiments (Table S4), the model performed very well in predicting the brand, with misclassifications only occurring at day 19. Similar high prediction performances were observed for Caution Wear, (Table S5a), Fantasy (Table S6a) and One (Table S7a) samples that were contaminated with dust. However, the model usually failed to enable identification of residues that were analyzed after transference to glass slides (Tables S5b, S6b and S7b).

To better understand the influence on the lubricant mass spectra of time since deposition, exposure to open air and/or dust, and transference to glass, on the ability of the model to make accurate predictions, ANOVA-simultaneous component analysis (ASCA) was used [26]. The resulting scores plots (loading factors are not shown) are presented in Fig. S19 for the variables time and open-air exposure (Kimono condom) in rolled and unrolled condoms, and: (1) time and dust exposure (Figs. S20); and (2) time and sample transference to a glass slide (Fig. S21) for the Fantasy condom. The results revealed that the factors that most influence the ability of the model to classify the condoms were extended time since deposition and transfer of lubricant to another surface. On the other hand, contamination with dust or exposure to open air exhibited little effect. The impacts on the spectra of open air exposure, time since deposition, dust contamination and transference to the glass slide (relative to fresh condom residue controls) were ~1–2%, ~5–22%, ~1–3% and ~24–28% respectively. However, it should be noted that the magnitude of the effect was dependent on the brand of condom tested. Nevertheless, the following general trends emerged from the ASCA results: (1) there was considerable variation in the spectra between days 1 and 19 (Figs. S19 and S20), as evidenced by the deviations of the day-1 results from those of day-19. Examination of the corresponding loadings plots revealed that the m/z values 130 and 284 were responsible for this variability; (2) the loadings associated with the scores plots generated for the identification of samples that had been transferred from one surface to another (Fig. S21) showed that m/z 89, 135, and 158 were most affected; and (3) 11 of 14 identified "universal" condom markers belong to the group of variables that remain unaffected by external factors such as age, dust-contamination and transference from one surface to another.

4. Conclusions

Analysis by DART-HRMS of the residues of 110 condom types representative of 16 brands from around the globe, showed this mass spectrometric technique to be a viable method for rapid sampling to generate diagnostic chemical fingerprint signatures. No sample preparation steps were required and the samples could be analyzed directly, with the acquisition of spectra requiring only a few seconds per analysis. Inherent in the data were attributes that enabled the development of a robust chemometric approach for the successful attribution of condom brand to a given residue, with high performance accuracies. The application of both the SDA and PLS-DA methods enabled the extraction of key characteristics of the data, including a subset of m/z values that were universally present for all of the condoms that were analyzed. Also determined were m/z values unique to specific brands, and other m/z values that were present in most of the brands. The redundancy of the results determined by both the SDA and PLS-DA methods served as a means of confirming the findings. The detection of molecules that were present in all the condoms analyzed provides the opportunity for further investigations into: (1) whether they can potentially be used as universal condom markers; and (2) their lifetimes within biological matrices derived from fingerprints, or vaginal and seminal fluids, among others. Lastly, examination of the effects of time since deposition, the presence of contaminants, and the influence of transference of the residue from one surface to another, on the prediction ability of the model, showed that while the model worked well in predicting the identities of aged and dust-contaminated samples, it was not as effective in identifying condom residues that were transferred from one surface to another.

Acknowledgements

This work was supported in part by Award no. 2015-DN-BX-K057, awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this presentation are those of the authors and do not necessarily reflect those of the Department of Justice.

Competing interest statement

The authors declare no competing interests.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.talanta.2018.09.101.

References

[1] R.D. Blackledge, M. Vincenti, Identification of polydimethylsiloxane lubricant traces from latex condoms in cases of sexual assault, J. Forensic Sci. Soc. 34 (4) (1994) 245–256, https://doi.org/10.1016/S0015-7368(94)72928-5.
[2] R.D. Blackledge, Viscosity comparisons of polydimethylsiloxane lubricants in latex condom brands via Fourier self-deconvolution of their FT-IR spectra, J. Forensic Sci. 40 (3) (1995) 467–469, https://doi.org/10.1520/JFS13806J.
[3] G.P. Campbell, A.L. Gordon, Analysis of condom lubricants for forensic casework, J. Forensic Sci. 52 (3) (2007) 630–642, https://doi.org/10.1111/j.1556-4029.2007.00411.x.
[4] J. Wolfe, D.L. Exline, Characterization of condom lubricant components using Raman spectroscopy and Raman chemical imaging, J. Forensic Sci. 48 (5) (2003) 1065–1074.
[5] T. Coyle, N. Anwar, A novel approach to condom lubricant analysis: in-situ analysis of swabs by FT-Raman spectroscopy and its effects on DNA analysis, Sci. Justice 49 (1) (2009) 32–40, https://doi.org/10.1016/j.scijus.2008.04.003.
[6] R. Bradshaw, R. Wolstenholme, R.D. Blackledge, M.R. Clench, L.S. Ferguson, S. Francese, A novel matrix-assisted laser desorption/ionisation mass spectrometry imaging based methodology for the identification of sexual assault suspects, Rapid Commun. Mass Spectrom. 25 (3) (2011) 415–422, https://doi.org/10.1002/rcm.4858.
[7] F. Burger, M. Dawson, C. Roux, P. Maynard, P. Doble, P. Kirkbride, Forensic analysis of condom and personal lubricants by capillary electrophoresis, Talanta 67 (2) (2005) 368–376, https://doi.org/10.1016/j.talanta.2005.03.040.
[8] P. Maynard, K. Allwell, C. Roux, M. Dawson, D. Royds, A protocol for the forensic analysis of condom and personal lubricants found in sexual assault cases, Forensic Sci. Int. 124 (2–3) (2001) 140–156, https://doi.org/10.1016/S0379-0738(01)00588-6.
[9] G.S. Lee, K.M. Brinch, K. Kannangara, M. Dawson, M.A. Wilson, A methodology based on NMR spectroscopy for the forensic analysis of condoms, J. Forensic Sci. 46 (4) (2001) 808–821, https://doi.org/10.1520/JFS15052J.
[10] T.P. Hollenbeck, G. Siuzdak, R.D. Blackledge, Electrospray and MALDI mass spectrometry in the identification of spermicides in criminal investigations, J. Forensic Sci. 44 (4) (1999) 783–788, https://doi.org/10.1520/JFS14553J.
[11] R.A. Musah, A.L. Vuong, C. Henck, J.R. Shepard, Detection of the spermicide nonoxynol-9 via GC-MS, J. Am. Soc. Mass Spectrom. 23 (5) (2012) 996–999, https://doi.org/10.1007/s13361-012-0353-7.
[12] R.A. Musah, R.B. Cody, A.J. Dane, A.L. Vuong, J.R. Shepard, Direct analysis in real time mass spectrometry for analysis of sexual assault evidence, Rapid Commun. Mass Spectrom. 26 (9) (2012) 1039–1046, https://doi.org/10.1002/rcm.6198.
[13] M. Maric, L. Harvey, M. Tomcsak, A. Solano, C. Bridge, Chemical discrimination of lubricant marketing types using direct analysis in real time time-of-flight mass spectrometry, Rapid Commun. Mass Spectrom. 31 (12) (2017) 1014–1022, https://doi.org/10.1002/rcm.7876.
[14] M.F. Mirabelli, D.R. Ifa, G. Sindona, A. Tagarelli, Analysis of sexual assault evidence: statistical classification of condoms by ambient mass spectrometry, J. Mass Spectrom. 50 (5) (2015) 749–755, https://doi.org/10.1002/jms.3584.
[15] S.E. Spencer, S.Y. Kim, S.B. Kim, K.A. Schug, Matrix-assisted laser desorption/ionization-time of flight-mass spectrometry profiling of trace constituents of condom lubricants in the presence of biological fluids, Forensic Sci. Int. 207 (1–3) (2011) 19–26, https://doi.org/10.1016/j.forsciint.2010.08.010.

[16] D. Ballabio, V. Consonni, Classification tools in chemistry. Part 1: linear models. PLS-DA, Anal. Methods 5 (16) (2013) 3790–3798, https://doi.org/10.1039/C3AY40582F.
[17] K. Sjöstrand, L. Clemmensen, R. Larsen, B. Ersbøll, G. Einarsson, SpaSM-a Matlab toolbox for sparse statistical modeling, J. Stat. Softw. 84 (2018) 1–37.
[18] L. Clemmensen, T. Hastie, D. Witten, B. Ersbøll, Sparse Discriminant Analysis, Technometrics 53 (4) (2011) 406–413, https://doi.org/10.1198/TECH.2011.08118.
[19] P. Hanchuan, L. Fuhui, C. Ding, Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy, IEEE Trans. Pattern Anal. Mach. Intell. 27 (8) (2005) 1226–1238, https://doi.org/10.1109/TPAMI.2005.159.
[20] V. Vovk, A. Gammerman, G. Shafer, Algorithmic Learning in A Random World, Springer, New York, 2005.
[21] G. Shafer, V. Vovk, A tutorial on conformal prediction, J. Mach. Learn Res. 9 (2008) 371–421.
[22] S. Wold, E. Johansson, M. Cocchi, PLS - Partial Least-squares projections to latent structures, in: H. Kubinyi (Ed.), 3D QSAR in Drug Design: Theory, Methods, and Applications, ESCOM, Leiden, The Netherlands, 1993, pp. 523–550.
[23] Flavor. ⟨https://en.wikipedia.org/wiki/Flavor⟩. (Accessed 21 June 2018).
[24] Aroma Compound. ⟨https://en.wikipedia.org/wiki/Aroma_compound⟩. (Accessed 21 June 2018).
[25] B. Cravatt, O. Prospero-Garcia, G. Siuzdak, N. Gilula, S. Henriksen, D. Boger, R. Lerner, Chemical characterization of a family of brain lipids that induce sleep, Science 268 (5216) (1995) 1506–1509, https://doi.org/10.1126/science.7770779.
[26] A.K. Smilde, J.J. Jansen, H.C.J. Hoefsloot, R.-J.A.N. Lamers, J. Greef, M.E. Timmerman, ANOVA-simultaneous component analysis (ASCA): a new tool for analyzing designed metabolomics data, Bioinformatics 21 (2005) 3043–3048, https://doi.org/10.1093/bioinformatics/bti476.