Comparison of Exhaled Endogenous Particles from Smokers and Non-Smokers Using Multivariate Analysis

Anna Bredberg
Mats Josefson
Ann-Charlotte Almstrand
Jukka Lausmaa
Peter Sjövall
Anna Levinsson
Per Larsson
Anna-Carin Olin

Occupational and Environmental Medicine, Sahlgrenska Academy at University of Gothenburg, Gothenburg, Analytical Science, Pharmaceutical Development, AstraZeneca R&D, Mölndal, and Chemistry and Materials Technology, SP Technical Research Institute of Sweden, Borås, Sweden

Key Words
Cigarette smoking · Inflammation · Orthogonal partial least squares · Particles in exhaled air · Phospholipids · Respiratory tract lining fluid · Time-of-flight secondary ion mass spectrometry

Abstract
Background: Smoking, along with many respiratory diseases, has been shown to induce airway inflammation and alter the composition of the respiratory tract lining fluid (RTLF). We have previously shown that the phospholipid and protein composition of particles in exhaled air (PEx) reflects that of RTLF. In this study, we hypothesized that the composition of PEx differs between smokers and non-smokers, reflecting inflammation in the airways. Objective: It was the aim of this study to identify differences in the phospholipid composition of PEx from smokers and non-smokers. Methods: PEx from 12 smokers and 13 non-smokers was collected using a system developed in-house. PEx was analysed using time-of-flight secondary ion mass spectrometry, and the mass spectral data were evaluated using multivariate analysis. Orthogonal partial least squares (OPLS) was used to relate smoking status, lung function and pack years to the chemical composition of RTLF. The discriminating ions identified by OPLS were then used as explanatory variables in traditional regression analysis. Results: There was a clear discrimination between smokers and non-smokers according to the chemical composition, where phospholipids from smokers were protonated and sodiated to a larger extent. Poor lung function showed a strong association with higher response from all molecular phosphatidylcholine species in the samples. Furthermore, the accumulated amount of tobacco consumed was associated with variations in mass spectra, indicating a dose-response relationship. Conclusion: The chemical composition of PEx differs between smokers and non-smokers, reflecting differences in the RTLF. The results from this study may suggest that the composition of RTLF is affected by smoking and may be of importance for lung function.

Introduction
Cigarette smoke contains more than 4,000 substances and the immune response to some of the harmful agents initiates and propagates chronic inflammation of the lungs.
of susceptible individuals [1, 2]. Inflammatory changes in airway disease may be reflected by alterations in the respiratory tract lining fluid (RTLF), whose main constituents are lipids and proteins [3–7]. Previous studies of RTLF using bronchoalveolar lavage (BAL) have shown differences in lipid and protein content in the RTLF of smokers and non-smokers. The smokers had increased levels of for example phosphoethanolamine, phosphatidylglycerol, annexin and glutathione S-transferase and lower levels of cholesterol, secretory IgA and surfactant protein A and D compared to non-smokers [8–10].

A novel method for monitoring the airways has been developed at the Department of Occupational and Environmental Medicine, University of Gothenburg, Gothenburg, Sweden. This non-invasive method involves the collection of endogenous particles in exhaled air (PEx) using a three-stage impactor and subsequent chemical analysis [11]. PEx originates from the RTLF, which covers the airways and acts as a protecting interface between the external environment and epithelial cells. The formation of PEx has not been fully elucidated but a significant mechanism of particle formation is airway re-opening after closure in the very distal bronchioles [12]. We have previously shown, using the same breathing and sampling technique as in the present study, that the lipid and protein profile of PEx largely resembles the profile observed in BAL [11, 13].

In the present study, extensive chemical analyses have been performed on PEx, using a sensitive mass spectroscopic method, time-of-flight secondary ion mass spectrometry (TOF-SIMS) [14, 15]. One of the advantages of TOF-SIMS analysis is that very small amounts of a sample are required and no sample preparation is needed. TOF-SIMS analyses of biological samples result in information-rich mass spectral data, with hundreds of peaks from molecular and fragment ions, which is a challenge to evaluate [16, 17]. Consequently, multivariate analysis, which allows the maximum amount of information to be extracted, has proven to be a valuable aid for the interpretation of TOF-SIMS data [18, 19]. The multivariate tool orthogonal partial least squares (OPLS) is a hypothesis-free statistical method well suited for the type of data in the present study, where the a priori knowledge on changes in phospholipid composition due to smoking is strictly limited [20]. The findings were then tested ad hoc with a traditional statistical approach.

The overall aim is to investigate whether PEx can be used to detect inflammatory changes in the RTLF. The specific aim of this study was to identify: (1) differences in the phospholipid composition of PEx from smokers and non-smokers and to determine the association between phospholipids, (2) the amount of tobacco consumed, and (3) the lung function, using TOF-SIMS, OPLS and conventional statistical methods.

**Methods**

**Study Participants**

The study included current smokers (n = 12) and non-smokers (n = 13), 49–86 years of age, who volunteered for the examination. Participants were included in the study if they were healthy, above 40 years of age and smokers or non-smokers. The smokers had smoked daily for at least 10 years and the non-smokers also included 1 ex-smoker who had quit smoking 20 years prior to the examination, the rest were never-smokers. The smokers abstained from smoking 1 h before the examination.

The study was approved by the Ethics Committee of the University of Gothenburg and all participants gave their written informed consent.

**Clinical Examination**

The fraction of exhaled nitric oxide (FENO) was measured using a NIOX Mino device (Aerocrene AB, Solna, Sweden) according to standard methodology, but only one measurement was performed for each subject, at an exhalation flow of 45–55 ml/s [21]. Spirometry was performed according to American Thoracic Society/European Respiratory Society standards [22] using a Spirare system (Diagnostica, Oslo, Norway) and related to reference values [23]. For all participants, PEx was collected after measurement of FENO and spirometry.

Allergic sensitization was defined as a positive skin prick test for common inhalant allergens or a positive Phadiatop test (Phadia, Uppsala, Sweden; for details, see online suppl. material; for all online suppl. material, see www.karger.com/doi/10.1159/000350941).

A questionnaire regarding smoking habits and respiratory condition, e.g., asthma, allergies, upper respiratory tract infections, chronic obstructive pulmonary disease, and medication, was filled in by all participants.

**Sampling of PEx**

A sampling device for the non-invasive collection of non-volatile compounds in exhaled air has been developed at the Department of Occupational and Environmental Medicine, University of Gothenburg, Gothenburg, Sweden [11]. Each subject rinsed his/her mouth with purified water and breathed particle-free air for 3 min prior to sampling. All subjects performed thirty forced exhalations into the sampling device. The device is equipped with a valve to divert the airflow when sampling, making it possible to breathe particle-free air tidally between the forced exhalations without sampling. The exhaled air containing PEx was split, with 8% (1.2 liters·min⁻¹) into a Grimm 1.108 optical particle counter and with 92% (13.8 liters·min⁻¹) into a 3-stage inertial impactor/pump (3-stage PM 10 Impactor, Dekati Ltd., Tampere, Finland) making it possible to simultaneously count, measure according to size and collect PEx by impaction on silicon plates. The particles are concentrated into ten separate spots on each plate. The silicon plates were stored at −20°C for no longer than 6 weeks, prior to analysis by TOF-SIMS.
Analysis of PEx with TOF-SIMS

TOF-SIMS analysis was performed using a TOF-SIMS IV instrument (ION-TOF, Münster, Germany) using a Bi³⁺ primary ion beam. The primary ion beam produced secondary ions from the analysed material which, in positive ion mode, comprised protonated ions (M + H)⁺ and cations containing alkali metal ions, e.g. sodiated (M + Na)⁺ and potassiated (M + K)⁺ ions, and in negative ion mode, comprised deprotonated ions (M – H). TOF-SIMS spectra were acquired on two randomly selected sample spots for each subject as previously described by Almstrand et al. [11]. The acquired spectra were combined into one average mass spectrum for each spot.

Statistical Analysis

Hypothesis-generating multivariate techniques, principal component analysis (PCA) and OPLS were applied for the statistical analysis. Peak intensities in the mass to charge (m/z) range of 20–1,000 were normalized to total ion intensity in the same range using a digital resolution of 0.2 m/z. The normalized peak intensities were imported into multivariate software (SIMCA software version 12, Umetrics AB, Umeå, Sweden) and Pareto scaled multiplicative (SIMCA software version 12, Umetrics AB, Umeå, Sweden) and Pareto scaled. Peak intensities had to be categorized into quartiles. The normalized peak intensities were imported into multivariate software (SIMCA software version 12, Umetrics AB, Umeå, Sweden) and Pareto scaled. PCA is an unsupervised method to examine the main variation (not necessarily related to the variable of interest) in the mass spectra and to identify possible gross outliers. The supervised method OPLS was applied in three models to compare variation in the selected responses (smoking status, pack years, lung function) and identifies possible gross outliers. The supervised method OPLS was applied in three models to compare variation in the selected responses (smoking status, pack years, lung function) and to identify possible gross outliers. The supervised method OPLS was applied in three models to compare variation in the selected responses (smoking status, pack years, lung function) and to identify possible gross outliers.

Results

The characteristics of the participants are summarized in Table 1. All participants were Caucasians and free of respiratory symptoms. No significant difference was observed between the groups with regard to age, sex, number of particles collected, body mass index or FEV₁/FVC. However, smokers had significantly lower FEV₁ pred and lower FENO, as expected. Two of the smokers had a FEV₁/FVC <0.7 and a FEV₁ pred between 50 and 79%. However, none of the participants had a previous diagnosis of chronic obstructive pulmonary disease and no reversibility test was performed in this study to confirm the diagnosis. All sampling and chemical analysis of PEx, in this study, was performed from April 2008 to March 2009.

Table 1. Summary of study participant characteristics

<table>
<thead>
<tr>
<th></th>
<th>Non-smoker</th>
<th>Smoker</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean age, years</td>
<td>64 (50–86)</td>
<td>62 (49–79)</td>
<td>0.71</td>
</tr>
<tr>
<td>Sex ratio, M/F</td>
<td>6/7</td>
<td>6/6</td>
<td></td>
</tr>
<tr>
<td>Particle amount</td>
<td>210,000 (28,000–410,000)</td>
<td>220,000 (9,500–1,600,000)</td>
<td>0.77</td>
</tr>
<tr>
<td>Body mass index</td>
<td>25 (22–29)</td>
<td>27 (18–34)</td>
<td>0.14</td>
</tr>
<tr>
<td>FEV₁ pred</td>
<td>105 (88–120)</td>
<td>91 (71–110)</td>
<td>0.016</td>
</tr>
<tr>
<td>FEV₁/FVC</td>
<td>0.76 (0.65–0.88)</td>
<td>0.71 (0.61–0.8)</td>
<td>0.094</td>
</tr>
<tr>
<td>NO, ppb</td>
<td>21 (10–102)</td>
<td>14 (3–23)</td>
<td>0.0069</td>
</tr>
<tr>
<td>Pack years</td>
<td>n/a</td>
<td>29.5 (10.5–51)</td>
<td></td>
</tr>
</tbody>
</table>

Age is presented as the mean, and all other parameters are summarized as the median, with ranges in parentheses. n/a = Not assessed.
The amount of PEx collected on the silicon plate varied substantially between individuals, ranging between 10,000 and 1,600,000 (median 220,000) for smokers and between 28,000 and 410,000 (median 210,000) for non-smokers, but did not differ significantly between the groups.

The number of PEx collected was not related to smoking status, age, body mass index or lung function, i.e. FEV$_1$, FEV$_1$ pred, FVC and FEV$_1$/FVC.

### TOF-SIMS Analysis

The main identified ions corresponding to phospholipids are summarized in Table 2. The molecular species of phospholipids are abbreviated as x:a, where x is the number of carbon atoms and a is the number of double bonds in the hydrocarbon chain. The most abundant positive molecular ions from phospholipids were phosphatidylcholine (PC), PC 30:0, PC 32:0 and PC 34:1, whilst the predominant negative molecular ions from phospholipids were phosphatidic acid (PA), PA 30:0 and PA 34:1, phosphatidyl glycerol (PG), PG 30:0, PG 34:1 and PG 36:1, and phosphatidyl inositol (PI), PI 34:1 and PI 36:1. However, in many negative ion spectra the intensities of molecular ions were below the detection limit and, therefore, these data were not included in statistical analysis.

### Multivariate Statistical Analysis

Prior to the multivariate evaluation of mass spectral data, ion abundances were normalized to the total intensity of the corresponding mass spectrum (mass range 20–1,000 m/z). No outliers were found in the PCA score plot, and all samples were included in the further processing of data. The PCA model used nine components with a predictive ability Q$^2$ = 0.86. From the PCA loading plot, it was concluded that the ions that varied most within the complete dataset were potassium (m/z 39.0) and the two PC fragments, m/z 86.2 and m/z 184.2. No separation between smokers and non-smokers was observed in the PCA model.

The association between smoker/non-smoker status and the mass-spectral variation was analysed using orthogonal partial least square-discriminant analysis (OPLS-DA). The resulting score plot showed a separation between the two groups according to positive ion mass spectra (Fig. 1). Each projected point represents the combined mass spectral information from one PEx spot.

The OPLS-DA model resulted in one predictive and four orthogonal components with the total predictive ability Q$^2$ = 0.55. The predictive component had an R$^2$X$\text{pred}$ of 0.029, meaning that approximately 3% of the variation in the data set was associated with the difference between smokers and non-smokers. The OPLS-DA analysis also provides loadings which contain the variation in m/z information to support the grouping of mass spectra in the score plot. The ions dominating for the smoker group represent a positive value in the loading plot while the prominent ions for the non-smoker group represent a negative value (Fig. 2). Investigation of the low m/z region indicated that PEx from smokers contained relative-

| Table 2. Assigned peaks from mass spectra and corresponding m/z ratios |
|---|---|---|---|
| **Positive ions** | **m/z** | **Negative ions** | **m/z** |
| PC 30:0 + H$^+$ | 706.5 | PA 32:0 – H | 647.5 |
| PC 30:0 + Na$^+$ | 728.5 | PA 34:1 – H | 673.5 |
| PC 32:0 + H$^+$ | 734.5 | PG 32:0 – H | 721.5 |
| PC 32:0 + K$^+$ | 744.5 | PG 34:1 – H | 747.5 |
| PC 32:0 + Na$^+$ | 756.5 | PG 36:1 – H | 775.5 |
| PC 34:1 + H$^+$ | 760.5 | PI 34:1 – H | 835.5 |
| PC 32:0 + K$^+$ | 772.5 | PI 36:1 – H | 863.5 |
| PC 34:1 + Na$^+$ | 782.5 | PI 36:1 + K$^+$ | 798.5 |
| PC 34:1 + Na$^+$ | 782.5 | PI 36:1 + K$^+$ | 798.5 |

The molecular species of phospholipids are abbreviated as x:a, where x is the number of carbon atoms and a is the number of double bonds in the hydrocarbon chain.
higher amounts of sodium and potassium and relatively lower amounts of phosphocholine compared to the non-smoker group (Fig. 3). The major difference in the high m/z region was associated with PC molecular ions. The smoker group samples contained lower levels of PC potassium adducts, corresponding to m/z 744.5, 772.5 and 798.5, compared to non-smokers. Furthermore, the smoker group samples contained greater levels of protonated and sodiated PC adducts, corresponding to m/z 706.5, 728.5, 734.5, 756.5 and 782.5, than the non-smoker group samples (Fig. 4). The ions most dominating in the S-plot from this model were m/z 23 and 39 (associated with high levels for smokers) and m/z 70 (high levels for non-smokers).
Secondly, the influence of total accumulated tobacco consumption on mass spectrum was investigated, and an association was observed between the number of pack years and the mass spectral variation. The resulting score plot from the OPLS model, with pack years as the continuous variable $Y$, showed that mass spectra from individuals with a higher tobacco consumption differed more from non-smokers than from individuals with a lower tobacco consumption ($R^2_{X\text{ pred}} = 0.025$ and $Q^2 = 0.62$) (fig. 5). The ions most important for the variation in the pack year model according to the S-plot were $m/z$ 39 and 46 (associated with a high number of pack years) and $m/z$ 70 and 81 (high levels for zero pack years).

The third chosen response, lung function ($FEV_1\text{ pred}$), was also correlated to the mass spectral variation ($R^2_{X\text{ pred}} = 0.10$ and $Q^2 = 0.63$). The results showed that low levels of lung function were associated with a high response from all molecular PC species present in the samples, with the exception of PC 34:1 $+\text{H}^+$. The ions most important for the variation in the lung function model according to the S-plot were $m/z$ 86.2 and 184.2 (associated with a high value for lung function) and $m/z$ 23 and 28 (associated with low lung function).

### Conventional Statistical Analysis

The main ions influencing the three models are listed in table 3 along with their p value.

<table>
<thead>
<tr>
<th>Model 1: smoking status</th>
<th>Response</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>23 ($\text{sodium, Na}$)</td>
<td>high for smoker</td>
<td>0.30</td>
</tr>
<tr>
<td>39 ($\text{potassium, K}$)</td>
<td>high for smoker</td>
<td>0.07</td>
</tr>
<tr>
<td>27</td>
<td>high for non-smoker</td>
<td>0.06</td>
</tr>
<tr>
<td>70</td>
<td>high for non-smoker</td>
<td>0.05</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model 2: amount of tobacco consumed</th>
<th>Response</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>39 ($\text{potassium, K}$)</td>
<td>high for many pack years</td>
<td>0.98</td>
</tr>
<tr>
<td>46</td>
<td>high for many pack years</td>
<td>0.02</td>
</tr>
<tr>
<td>70</td>
<td>high for few pack years</td>
<td>0.15</td>
</tr>
<tr>
<td>81</td>
<td>high for few pack years</td>
<td>0.11</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model 3: lung function</th>
<th>Response</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>86.2</td>
<td>high for high $FEV_1\text{ pred}$</td>
<td>0.76</td>
</tr>
<tr>
<td>184.2 ($\text{phosphocholine}$)</td>
<td>high for high $FEV_1\text{ pred}$</td>
<td>0.29</td>
</tr>
<tr>
<td>23 ($\text{sodium, Na}$)</td>
<td>high for low $FEV_1\text{ pred}$</td>
<td>0.44</td>
</tr>
<tr>
<td>28</td>
<td>high for low $FEV_1\text{ pred}$</td>
<td>0.03</td>
</tr>
</tbody>
</table>

1 Logistic regression model.
2 Linear regression model.

The interpretation of data was focused on PC species as they are the major constituent of surfactant, which in turn is an important part of the RTLF. It was apparent that the ionization of PC was different in smokers; smokers had more protonated and sodiated PC adducts compared to non-smokers. This is a novel finding, and the underlying mechanisms are yet unclear. A fraction of these molecular ions fragments into smaller ions as a nat-
ural consequence of the mass spectral analysis. The signals from these fragment ions also support the discrimination of smokers and non-smokers.

The results are strengthened by the fact that similar findings as in the smoker/non-smoker model were detected when the number of pack years was correlated to the mass spectral data. In the resulting score plot, the influence of tobacco smoke on mass spectrum was projected as a pack year gradient from left (non-smokers and smokers with the lowest number of pack years) to right (smokers with the highest number of pack years), displaying a dose-dependent change in the composition of RTLF.

Relating lung function to mass spectral data showed that mass spectra from individuals with low lung function were associated with a high response from all molecular PC species. In a previous study, we showed that individuals with asthma also had a higher response from molecular PC species compared to controls; however, there, the association to lung function was not analysed [28]. This may suggest that phospholipid composition of the RTLF is important for lung function.

The main results from the three models in the OPLS analysis were also tested using conventional statistical methods, and the results were overall confirmed.

One major limitation of this study was that the number of participants was small, and hence, the results must be interpreted with caution. The possibility that sodium or potassium from ambient air contaminated the samples during the collection procedure can be ruled out, since the analysis of reference spots (a sample-free area of the same silicon plate) for ten different samples revealed no relation with the amount of sodium or potassium in the samples. Additionally, to avoid any systemic bias, all samples were handled in the same manner and were collected and analysed in a randomized order.

A future possible improvement would be to control the number of particles collected as the variability of particle number was high between individuals and potentially influenced the mass spectral peak pattern [15]. However, neither lung function nor smoking status was related to the amount of particles produced. This was also controlled for in the conventional statistical analyses and the results were confirmed.

The TOF-SIMS analysis of small, non-volatile compounds is highly sensitive and is readily applicable to PEx because the sample is deposited onto silicon plates, which can be analysed directly without further sample preparation. In this study, thirty forced exhalations were collected but substantially less would have been sufficient to achieve the detection of phospholipids in PEx. Based on particle count, the number of exhalations can be individually optimized.

PEx has been demonstrated to be formed in the distal airways during airway opening that follows airway closure [12]. In the present study, dynamic compression during forced exhalation could be an additional mechanism of PEx formation, which would mean that the particles can be formed in more proximal airways. The lipid profiles observed in the present study and in a previous
study, using the same breathing and sampling techniques, were very similar to the lipid profile of BAL [11].

In conclusion, analysis of PEx is a new technique to monitor the chemical composition of RTLF in respiratory disease. The results indicate that chemical composition of PEx reflects changes in RTLF related to smoking and lung function. The results need further exploration in a larger study.

Acknowledgements

We gratefully acknowledge the dedicated contributions of Frederikke Johansson and Ingrid Wulcan, medical students at Gothenburg University, during the sampling of exhaled particles and the assistance provided by Per Borchartd, Chemist at SP Technical Research Institute of Sweden, with the TOF-SIMS analyses. We also thank Ekaterina Mirgorodskaya and Björn Bake for their fruitful discussion of the manuscript. The study was funded by the Swedish Heart-Lung Foundation (254212008) and Formas (254212006).

References


