BioMetals (2005) 18:603–613
DOI 10.1007/s10534-005-3001-x


# Proteomic profiling in the sera of workers occupationally exposed to arsenic and lead: identification of potential biomarkers

Rihong Zhai[1,2,*], Suhua Su[2], Xin Lu[3], Ruiqing Liao[2], Xianmin Ge[2], Min He[4], Yuanjiao Huang[2], Sui Mai[2], Xi Lu[2] & David Christiani[1,*]
[1]*Department of Environmental Health, Occupational Health Program, Harvard School of Public Health, 665 Huntington Ave., Boston, MA 02115, USA;* [2]*Guangxi Institute of Occupational Health, Nanning, 530021, China;* [3]*Department of Biostatistics, Harvard School of Public Health, 665 Huntington Ave., Boston, MA 02115, USA;* [4]*Central Laboratory, Guangxi Medical University, Nanning, 530021, China; *Author for correspondence (Tel: + 1-617-432-1641; Fax: + 1-617-432-6981; E-mail: rzhai@hsph.harvard.edu; dchris@hohp.harvard.edu)*

Received 18 July 2005; accepted 7 September 2005

*Key words:* biomarkers, proteomic profiles, metal mixture exposure, SELDI-TOF

**Abstract**

Arsenic (As) and lead (Pb) are important inorganic toxicants in the environment. Frequently, humans are exposed to the mixtures of As and Pb, but little is known about the expression of biomarkers resulting from such mixed exposures. In this study, we analyzed serum proteomic profiles in a group of smelter workers with the aim of identifying protein biomarkers of mixed As and Pb exposure. Forty-six male workers co-exposed to As and Pb were studied. Forty-five age-matched male office workers were chosen as controls. Urine As and blood Pb concentrations were determined. Serum proteomic profiles were analyzed by Surface-Enhanced Laser Desorption/Ionization Time-Of-Flight (SELDI-TOF) mass spectrometer on the WCX2 ProteinChip. Using Recursive support vector machine (RSVM) algorithm, a panel of five peptides/proteins (2097 Da, 2953 Da, 3941 Da, 5338 Da, and 5639 Da) was selected based on their collective contribution to the optional separation between higher metal mixture exposure and non-exposure controls. Among these five selected markers, the 3941Da was down-regulated and the four other proteins were up-regulated. Descriptive statistics confirmed that these five proteins differed significantly between metal exposure and non-exposure. Interestingly, the combined use of the five selected biomarkers could achieve higher discriminative power than single marker. These results demonstrated that proteomic technology, in conjunction with bioinformatics tools, could facilitate the discovery of new and better biomarkers of mixed metal exposure.

## Introduction

The health effects of chronic exposure to arsenic (As) and lead (Pb) have been extensively studied and well documented. Both are among the US EPA Superfund top ten priority hazardous substances (De Rosa *et al.* 1996). As is classified as a human carcinogen by International Agency for Research on Cancer (IARC) (Gebel 2000; Hughes 2002; IARC 2003). Chronic As exposure has been implicated in several non-cancerous conditions, in particular, skin disease, diabetes mellitus, hypertension and cardiovascular disease, perturbed porphyrin metabolism, hyperpigmentation, hyperkeratosis and irreversible noncirrhotic portal hypertension (NRC, 2001). It has been long known that As exposure is associated with skin, bladder and lung cancer (WHO 2001). In the majority of cases in which an internal cancer has been ascribed to As exposure, a dermatologic hallmark of As poisoning was also identified (Tsai *et al.* 1999). Pb is one of the oldest known and more widely studied

*Table 1.* General characteristics of the investigated groups.

| | Metal workers | Controls |
|---|---|---|
| No. | 46 | 45 |
| Sex | Male | Male |
| No. of smokers[a] | 24 | 22 |
| No. of nonsmokers | 22 | 23 |
| No. of alcohol drinker[b] | 28 | 28 |
| Age, years (±SD) | 39.9 ± 8.1 | 40.6 ± 8.5 |
| Years of metal exposure | 14.8 ± 10.2 | 0 |
| As levels in urine (μg/ml) | 0.22 ± 0.09 | 0.07 ± 0.03 |
| Pb levels in blood (μg/l) | 410.31 ± 106.88 | 83.98 ± 27.65 |
| Mean air concentration of Pb at work places ($mg/m^3$)[c] | 1.67 (0.18–11.31) | <0.004[d] |
| Mean air concentration of As at work places ($mg/m^3$)[c] | 0.56 (0.11–0.71) | <0.018[d] |

[a]Subjects smoked ≥1 cigarette/day were defined as smokers.
[b]Individuals took alcohol ≥100 ml/week were considered as alcohol drinker.
[c]The Hygienic Standards in China are 0.05 $mg/m^3$ for Pb and 0.02 $mg/m^3$ for As.
[d]The detectable limit of the analytical methods (CNCC, 1995, 1997).

toxicant and evidence of Pb poisoning can be found dating back to Roman times. Pb can induce multiple adverse effects including reproductive toxicity, neurotoxicity, carcinogenicity, nephrotoxicity, immunotoxicity, and hypertension (Gidlow 2004). In occupationally unexposed adult populations, blood Pb levels ≤250 μg/l were generally considered allowable (WHO 1995). However, neurotoxic effects have been detected in children with blood Pb level <100 μg/l, suggesting that safe Pb levels remain unclear (Koller *et al*. 2004).

In environmental and occupational settings, humans are often exposed to combinations of metals. For instance, As and Pb most often occur together and are present in the top ten binary combinations of contaminants in soil and water (Fay & Mumtaz 1996); major exposures in metal smelters would have included As and Pb (Lilis *et al*. 1985; Binks *et al*. 2005). Although different metals have unique primary mechanisms of action that are cell and/or tissue specific (Goyer 1996), studies have suggested that metals can interact one another's cytotoxicity and/or carcinogenic potentials (Nordberg & Anderson 1981; Elsenhans *et al*. 1987; Bae *et al*. 2001). As a result, we still lack a fundamental understanding of the actions of metal mixtures at the molecular level. Accurate risk assessment of these highly relevant chemicals awaits progress in this area, including biomarkers of preclinical disturbances following exposure. However, much of the effort in the past has centered largely on biomarkers of a single hazard. Examples include As levels in urine and nails as biomarkers of As exposure (WHO 2003), Pb concentrations in blood/urine/plasma, delta- aminolevulinic acid (ALA) in urine/blood/plasma; coproporphyrin in urine (CP), and zinc protoporphyrin (ZPP) in blood as biomarkers of Pb exposure and effect (Sakai 2000). Nevertheless, little progress has been made in identifying biomarkers of metal mixture exposure.

Proteomic analysis is a powerful technology, developed recently to enhance research on the diagnosis, treatment and prevention of human diseases (Röcken *et al*. 2004). By examining comprehensively different protein profiles (expression level, post-translational modification, and interaction) between normal and diseased or drug-treated samples, proteomics may provide information on new biomarkers, disease-associated targets and the process of pathogenesis (Xiao *et al*. 2005). This technique has been used successfully to identify specific serum proteomic biomarkers that could discriminate cancer from non-cancer in bladder (Vlahou *et al*. 2001), lung (Zhukov *et al*. 2003), breast (Li *et al*. 2002), and ovarian cancers (Petricoin *et al*. 2002).

Because proteins are gene products, it is logical to expect that specific serum proteomic patterns might reflect an underlying pathological state in humans. The objective of the present study was to apply proteomic technology to explore potential proteomic biomarkers in workers exposed to As and Pb mixture. Our results showed that proteomic methods, in combination

with bioinformatics, could lead to identification of new biomarkers that differentiate subjects exposed to high metal mixture from non-occupationally exposed individuals.

## Materials and methods

### *Study population*

The study group consisted of 46 male workers actively employed in a metal smelter in Guangxi, China. This refinery extracts antimony and tin from tin mine produced by Dachang Tin mine Co., a mine previously known for high prevalence of lung cancer and high content of As in the raw mines (Chen & Chen 2002). Environmental monitoring in this smelter showed that the mean air concentrations of Pb and As at workplaces were higher than the National Hygienic Standards of China (Ministry of Health, China, 2002). Subjects in this study were workers who had been found to have either higher level of blood Pb or urine As, in routine occupational health surveillance. In this group, 74% (34/46) had blood Pb >400 $\mu$g/l (a value used for the diagnosis of occupational chronic Pb poisoning according to China criteria; CNCC 2002), and 46% (21/46) had urine As level >0.24 $\mu$g/ml (the local community background urine As level, Li *et al.* 2004). The control group included 45 male office employees who had no previous history of occupational exposure to metals. No As- or Pb-containing material has been handled in their office, and the air concentrations of As and Pb in their office were undetectable. The general characteristics of the population are shown in Table 1.

Protocols for this study were reviewed and approved by the Institutional Review Board (IRB) of Guangxi Institute of Occupational Health, China. After giving informed consent, each subject completed a questionnaire to provide information on demographics, occupational and medical histories. All subjects underwent a medical examination including physical exam, chest X-ray, liver function, blood and urine routine tests, ultrasonography of the liver and kidney. To obviate the influences of possible confounding factors, subjects with abnormal findings in liver function, chest X-ray, blood/urine routine, higher body temperature, hypertension, and those who used of medication in the month prior to this investigation were excluded from the study. For all subjects, blood and urine samples were collected in the morning of the same day before going to work. A blood sample was collected by venipuncture into tubes and spun at 2500 rpm at 4 °C for 10 min and then stored at −80 °C until analysis. Subjects were asked to wash their hands before supplying the urine, in order to reduce contamination. All urine sample containers were immersed in liquid nitric acid for 48 h to eliminate contaminating metals before they were used for collecting urine samples.

### *Measurements of blood Pb and urine As*

Blood lead (BPb) level was quantified using graphite furnace atomic absorption spectrophotometer (Sollaar M6, Thermo Electron Corp., USA) with the methods recommended by China National Standard Committee (CNCC 2002; comparable to protocol of NIOSH 1994). Urine As was determined using spectrophotometer according to the protocol of CNCC (1996; comparable to method of Mazur *et al.* 1983). To ensure the accuracy of analysis, quality control procedures were applied to all measurements. Briefly, a standard curve was generated using known concentrations of As or Pb. The reference standards of As and Pb were purchased from the National Research Center for Certified Reference Materials (NRCCRM), Beijing, China. Values for samples were determined using the linear portion of the curves for standards run in parallel with each batch of assays. In addition, a sample of blind reference material was included in each batch of analysis. When the value of the reference material was within the expected range, the result of the sample was considered acceptable.

### *SELDI TOF analysis of serum protein profiles*

Protein profiling of the serum samples was performed using the eight-spot format WCX2 (weak cationic exchange) ProteinChip Arrays (Ciphergen Biosystems, Fremont, CA, USA). Frozen serum samples were thawed and spun at 10,000 rpm for 5 min at 4 °C. Twenty microliter of U9 buffer was added to 10 $\mu$l of each serum sample then shacked on ice for 30 min, before adding 360 $\mu$l WCX-2 buffer. Arrays were prepared as follows: each array was pre-equilibrated 2 × 5 min in 200 $\mu$l WCX-2 buffer on a horizontal shaker (MSI Minishaker)

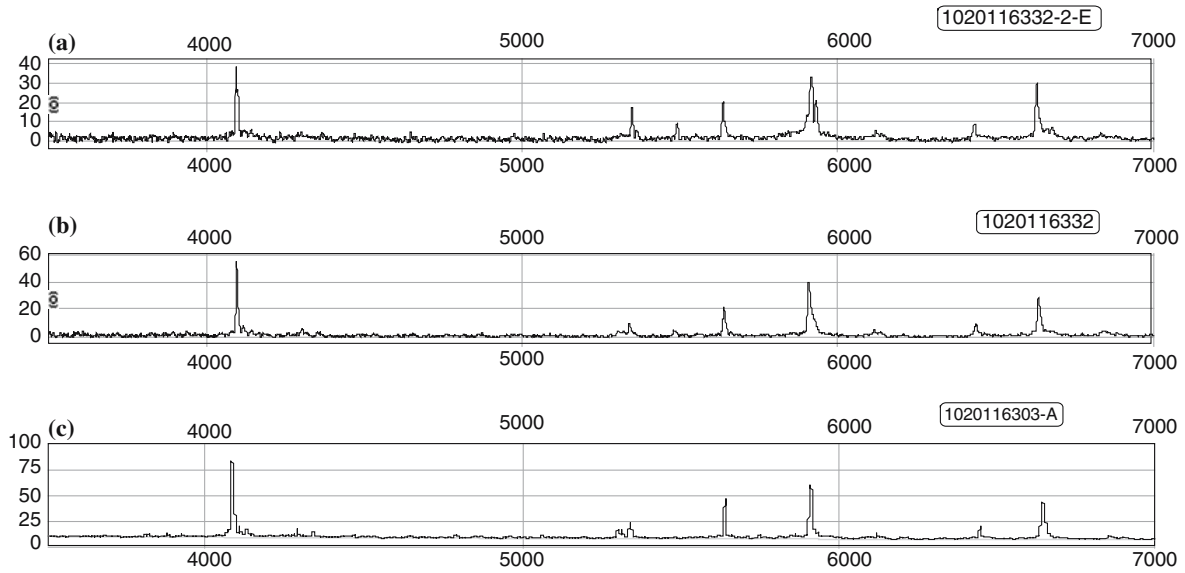


*Figure 1.* Reproducibility of proteomic spectra. Representative mass spectra from a serum specimen processed on the same chip array on the same day (a and b) and on a different chip analyzed one month later (c). X-axis, mass/charge ($m/z$) in Daltons; y-axis, relative intensity.

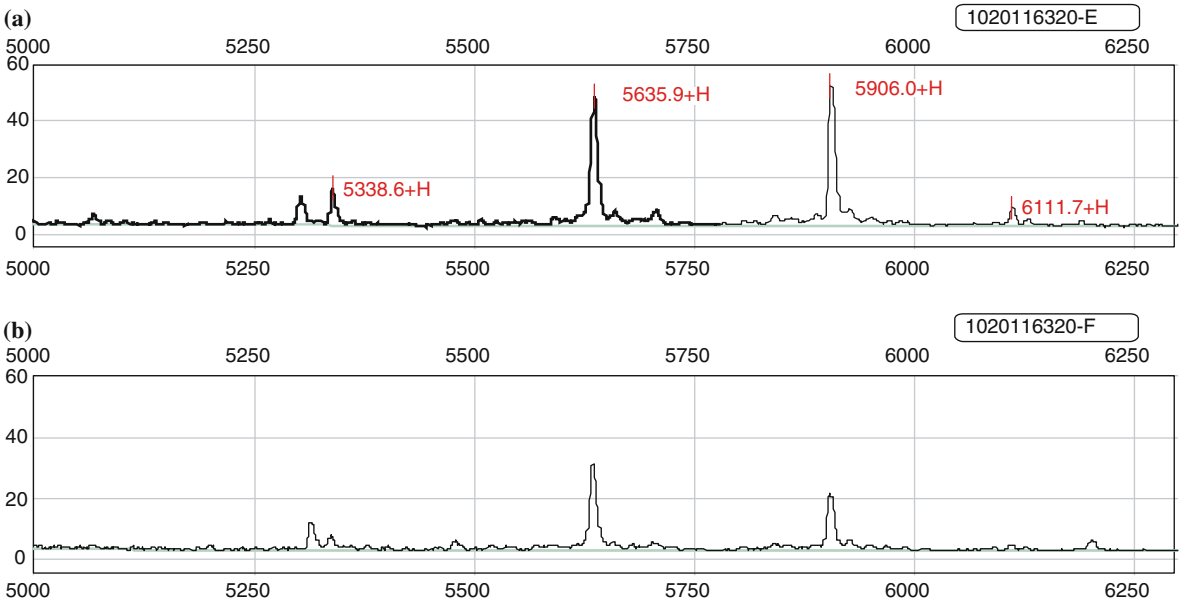


*Figure 2.* Representatives of serum protein profiles on WCX2 protein chips. The protein peaks of 5338 kDa, 5639 kDa, 5907 kDa, and 6116 kDa were up-regulated in the sample from metal worker (a) comparing to that from control (b). X-axis, mass/charge ($m/z$) in daltons; Y-axis, relative intensity.

before sample addition. Sample supernatant was added and incubated for 1 h on shaker. After incubation, the sample was removed, and each spot was washed with 200 $\mu$l WCX-2 buffer for 2 × 5 min with agitation. After washing, array was dismantled carefully from the bioprocessor and washed briefly with deionized water. Sinapinic acid (SPA) of 0.5 $\mu$l was deposited on the array spots

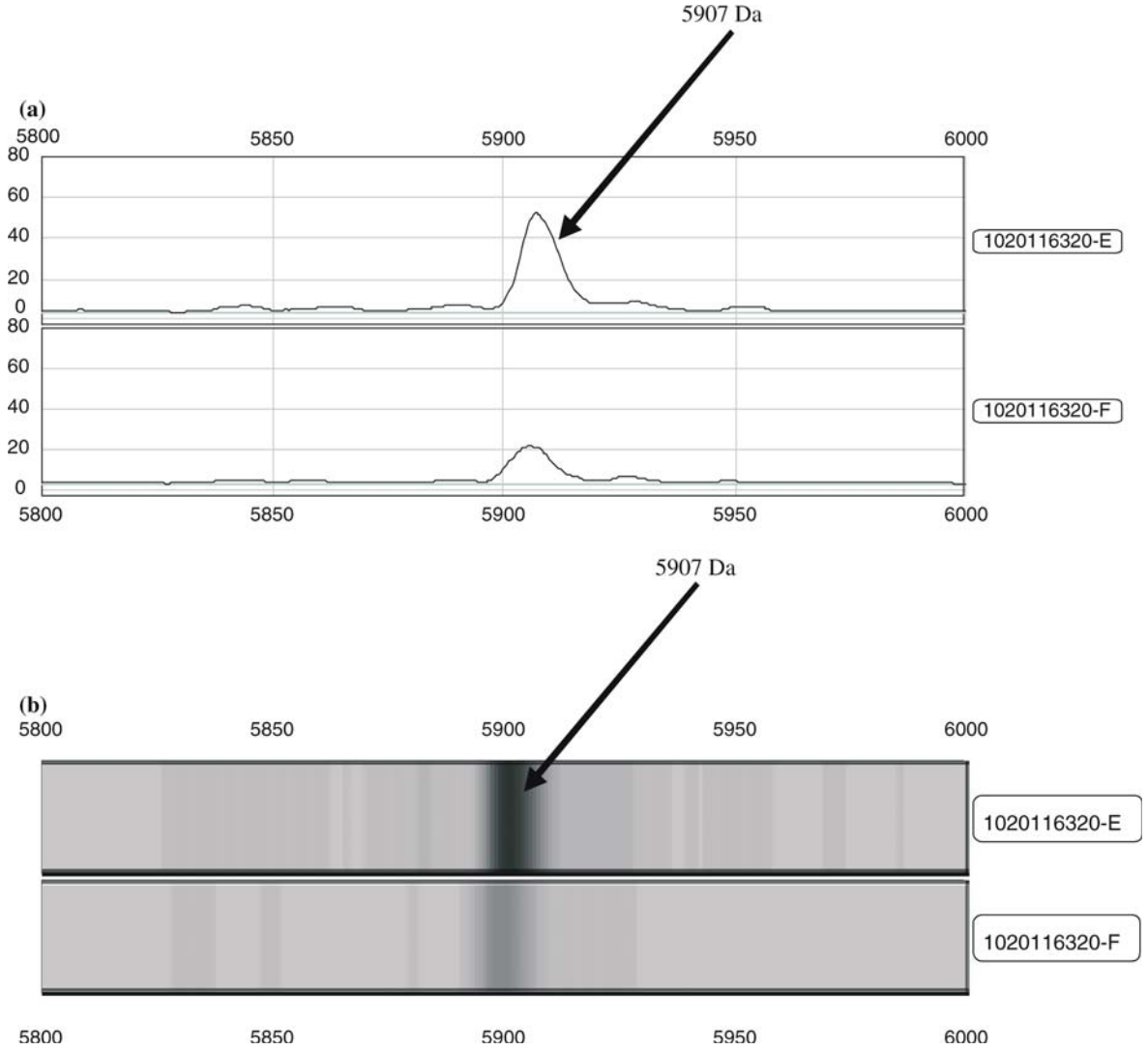


*Figure 3*. Representative spectra (a) and gel views (b) of the selected biomarkers. The serum samples from metal worker (-E) and control (-F) were run side by side on the same Proteinchip. The 5907-Da protein up-regulated in metal worker but down-regulated in control. X-axis, mass/charge ($m/z$) in daltons; Y-axis, relative intensity.

and allowed to air dry. The ProteinChip Arrays were read by surface-enhanced laser desorption/ionization time-of flight (SELDI-TOF) mass spectrometry (ProteinChip PBS II reader, Ciphergen), and time-of-flight spectra were generated by averaging laser shots with a laser intensity of 185, detector sensitivity of 8. Mass accuracy was calibrated externally using the All-in-1 peptide molecular weight standard (Ciphergen). Peak detection and alignment were performed with Biomarker Wizard function in Ciphergen ProteinChip Software V 3.0. Protein peak intensity was expressed as mass to charge ration ($m/z$). Peak intensities were normalized according to total ion current after background subtraction. For Biomarker Wizard setting, the signal-to-noise ratio was set between 2 and 5. Reproducibility was estimated using four representative serum samples: two from healthy controls and two from metal-exposed workers. Each representative serum sample was divided into three parts. Two parts were run in parallel on the same protein chip; the other part was tested one month later on different WCX2 chip under the same experimental conditions.

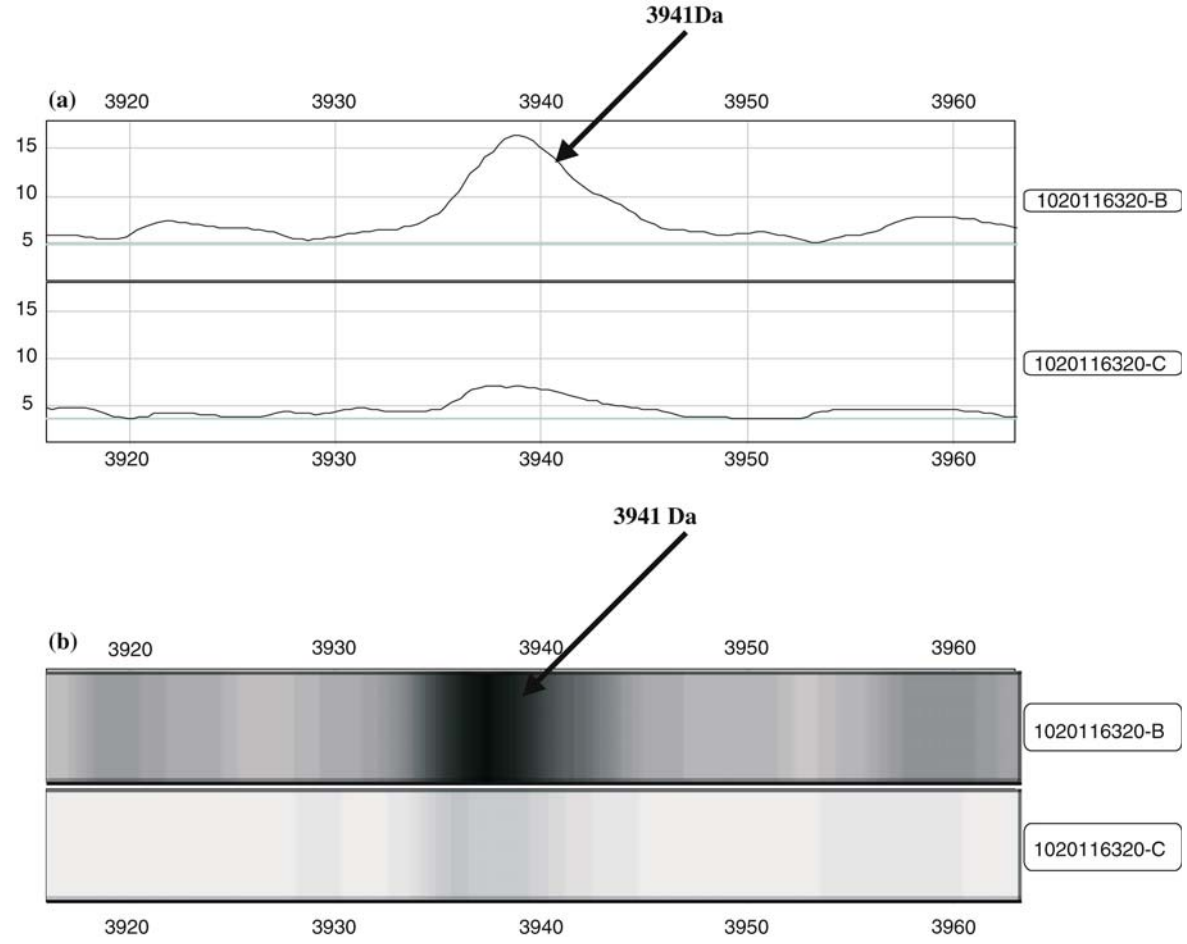


*Figure 4*. Representative spectra (a) and gel views (b) of the selected biomarkers. The serum samples from metal worker (-C) and control (-B) were run side by side on the same Proteinchip. The 3941-Da protein down-regulated in metal worker but up-regulated in control. X-axis, mass/charge ($m/z$) in daltons; Y-axis, relative intensity.

*Data analysis*

A linear recursive support vector machine (RSVM) algorithm was used to search for a combination of biomarkers that could segregate normal samples from highly exposed samples, and estimated the model prediction error at the same time (Xu *et al*. 2004). This feature selection algorithm was designed to avoid bias in the estimation of classification error. RSVM used the mean-weighted support vector machine (SVM) kernel to evaluate the relative importance of peaks, eliminates peaks iteratively by backward feature elimination process, and estimates the prediction error by unbiased external CV Scheme. Finally, this method generates a list of important peaks by frequency-based selection method (Xu *et al*. 2004). For details of the RSVM algorithm, please refer to: http://www.biostat.harvard.edu/~xzhang/R-SVM/R-SVM.html.

Comparison of the peak intensity between the controls and metal-exposed workers were performed using the Student *t*-test. The power of protein peak to discriminate subjects with higher metal exposure from healthy controls was estimated by the area under the curve (AUC) of receiver operating characteristic (ROC) (Eng J), which ranges from 0.5 (no discrimination) to 1.0 (absolute prediction) (Hanley & McNeil 1982). The peaks with an AUC <0.60 were excluded from further data analysis. Sensitivity was defined as the conditional probability of predicting metal

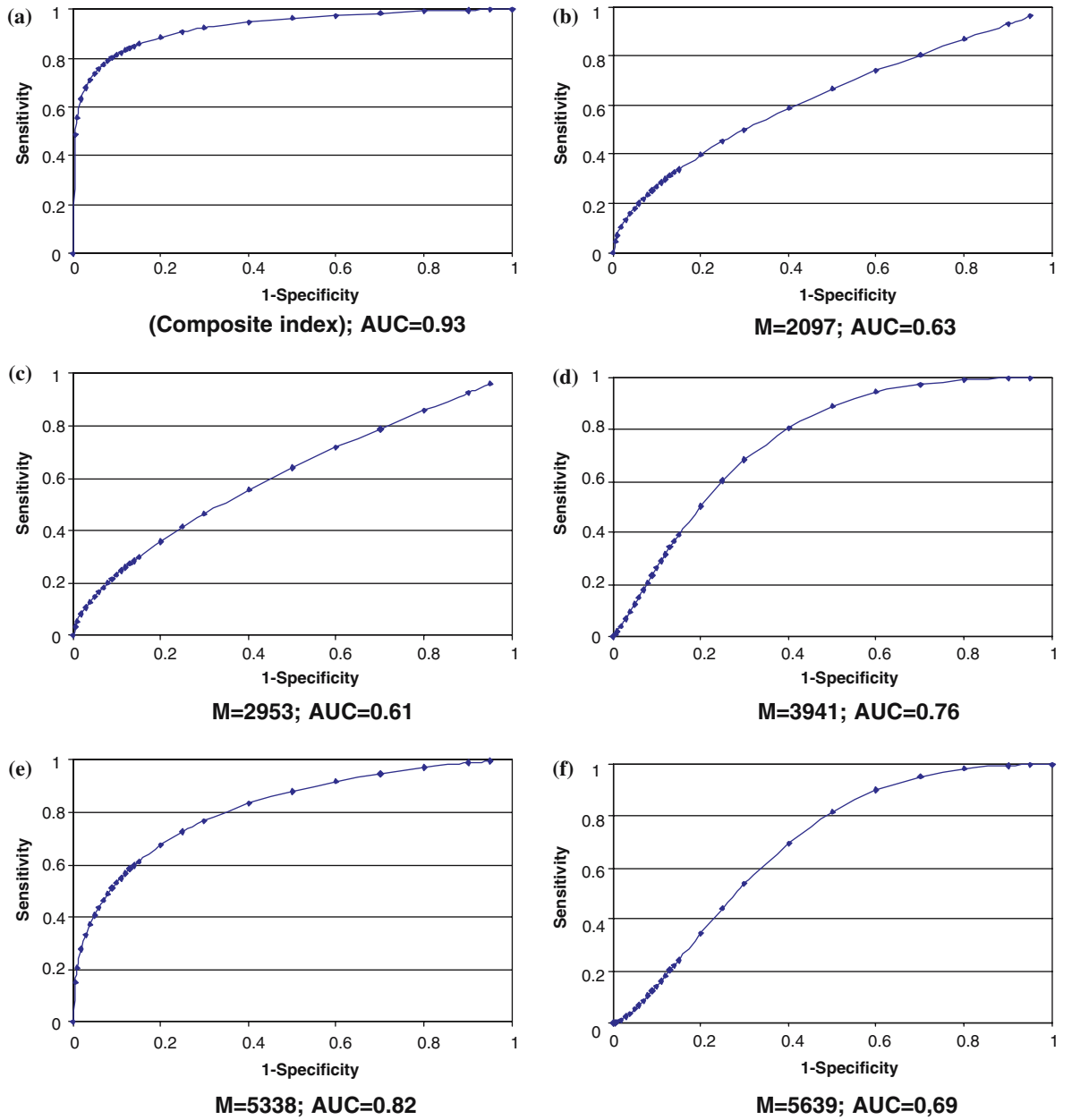


*Figure 5*. ROC curve analysis of the five RSVM selected protein markers (b, c, d, e, and f) and logistic-derived composite index (a); M, mass in daltons.

exposure given the gold standard is urine As >0.24 μg/ml or blood Pb >400 μg/l (the 95% upper limit values of the local population). Likewise, we defined specificity as the conditional probability of predicting non-metal exposure given that the gold standard as non-metal exposure (urine As <0.24 μg/ml or blood Pb < 400 μg/l).

## Results

### *Protein peak detection and data preprocessing*

Serum proteins retained on the WCX2 array were analyzed on a PBS II mass reader. The high mass to acquire was set to 50 kDa, with an optimization range from 1.5 to 15 kDa, because this range

*Table 2.* Serum proteomic peaks and mean values between controls and metal-exposed workers.

| Molecular mass ($m/z$) | Peak intensity (controls) $n=45$ | | Peak intensity (metal workers) $n=46$ | | $p$-value | AUC |
|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | | |
| 3941[a] | 5.29 | 3.67 | 2.97 | 2.82 | <0.0001 | 0.76 |
| 5338[a] | 4.28 | 2.31 | 8.71 | 5.66 | <0.000 | 0.82 |
| 5907 | 19.02 | 10.83 | 28.64 | 14.19 | 0.001 | 0.69 |
| 5639[a] | 12.00 | 5.52 | 15.72 | 6.45 | 0.002 | 0.69 |
| 5809 | 1.18 | 0.56 | 1.82 | 1.13 | 0.005 | 0.69 |
| 8136 | 1.82 | 1.02 | 1.32 | 0.73 | 0.007 | <0.61 |
| 2953[a] | 4.15 | 2.42 | 6.58 | 5.80 | 0.011 | 0.61 |
| 6116 | 2.15 | 1.13 | 3.13 | 1.91 | 0.013 | 0.66 |
| 5920 | 5.40 | 6.09 | 7.62 | 7.82 | 0.020 | 0.68 |
| 8183 | 1.19 | 0.76 | 0.92 | 0.64 | 0.020 | <0.61 |
| 15,305 | 0.40 | 0.26 | 0.83 | 0.96 | 0.021 | <0.61 |
| 1860 | 1.78 | 1.05 | 1.33 | 1.16 | 0.024 | <0.61 |
| 5855 | 2.35 | 2.76 | 2.64 | 1.86 | 0.025 | 0.65 |
| 3960 | 1.96 | 1.07 | 1.52 | 1.10 | 0.027 | <0.61 |
| 4103 | 8.56 | 6.36 | 5.84 | 4.54 | 0.030 | <0.61 |
| 11,812 | 0.43 | 0.15 | 0.55 | 0.28 | 0.031 | <0.61 |
| 1776 | 1.23 | 1.36 | 1.87 | 1.92 | 0.038 | <0.61 |
| 2097[a] | 1.42 | 1.29 | 2.42 | 2.51 | 0.043 | 0.63 |
| 2127 | 3.40 | 5.19 | 2.10 | 2.22 | 0.047 | <0.61 |
| 4299 | 4.87 | 2.78 | 3.78 | 2.31 | 0.048 | <0.61 |
| 10,264 | 0.83 | 0.48 | 0.65 | 0.46 | 0.048 | <0.61 |

[a]Biomarkers selected by RSVM.

contained the majority of the resolved protein/peptides. A mass accuracy of 0.1% was achieved by external calibration using the Standard (Ciphergen).

Among a total of 327 qualified mass peaks (signal-to-noise ratio > 5) detected, 256 peaks had $m/z$ values between 1.5 and 5 kDa, 58 peaks had $m/z$ values between 5 and 10 kDa, and 13 peaks were between 10 and 15 kDa. Peaks with a $m/z$ < 1.5 kDa were mainly ion noise from the matrix and therefore excluded. Peak intensity was normalized to total ion current, and logarithmic transformation was applied.

Figure 1 shows the reproducibility of three SELDI protein spectra obtained either from the same run on the same chip or from another run on another protein chip for the serum sample of the same subject. Figure 2 illustrates the comparison of proteomic profiles between a sample of metal worker and that of control. Protein patterns generated from both samples were virtually similar with regard to the peaks detected, but the relative mass intensities of some markers were different.

*Selection of significant biomarkers and classification of proteomic patterns using the RSVM algorithm*

To identify biomarkers with potential for discrimination of higher metal mixture exposure, a RSVM algorithm was used to select a panel of statistically significant biomarkers and to segregate sample classes. There were in total 37,223 data points in the whole spectrum. Since there was hardly any signal in the region above 20 kDa and too much noise below 1.5 kDa, only the region between 1.5 and 20 kDa was used in the following analysis, where there were 29,757 data points.

The biomarker detection function of Ciphergen software detects 327 biomarkers from the region between 1.5 and 20 kDa. From the comparison between metal-exposed worker and control samples, the RSVM feature selection algorithm identified five important protein markers: 2097 Da, 2953 Da, 3941 Da, 5338 Da, and 5639 Da.

*Biomarker discovery and identification using descriptive statistics*

The Student *t*-test was used to evaluate the differences of single peak intensity between metal workers and controls. Among 327 protein peaks, the intensities of 12 proteomic peaks were significantly higher and 9 significantly lower in the sera of metal workers than that in controls ($p$-value $<0.05$; Figures 3 and 4). Thus, 21 serologic proteomic peaks were statistically significant (Table 1), but not necessarily all at the highest ranking. Interestingly, the RSVM-selected five serum protein markers were all significantly different between metal workers and controls as calculated by *t*-test.

To estimate the power of the potential biomarkers, we performed receiver operating characteristic (ROC) analysis on 21 candidate peaks identified by *t*-test. In this analysis, ten protein peaks possessed an AUC value $\geq 0.60$ (Table 2). Again, the AUCs of RSVM-selected five protein peaks were among these ten top markers. However, the values of AUCs for individual marker were generally moderate, suggesting that the power of single protein marker to differentiate higher metal exposure to non-exposure was relatively low. To show the collective power of multiple biomarkers, multivariate logistic regression was used to combine the RSVM-selected five biomarkers to form a single-value composite index. ROC curve analysis of the composite index showed a much stronger AUC (0.93, sensitivity = 87%, specificity = 89%) compared with the AUCs from individual biomarkers (Figure 5). This result illustrated the power of the combination of multiple important markers.

## Discussion

In this study, we attempted to search for biomarkers of mixed As and Pb exposure, using a serum proteome-based approach. Using SELDI ProteinChip technology, we were able to analyze comprehensive protein profiles of subjects exposed to metal mixture. The machine-learning algorithm allows evaluation of each mass according to its collective contribution toward the maximal separation of mixed metal exposure at higher levels from non-exposure. The two methods led to the identification of five discriminatory protein peaks (2097 Da, 2953 Da, 3941 Da, 5338 Da, and 5639 Da). These five proteins formed a proteomic signature that, in combination, achieved higher sensitivity and specificity in detecting higher metal mixture exposure from non-exposure than single protein marker.

Biomarker discovery using proteomics not only involves the simultaneous analysis of proteomic profiling, but also sophisticated bioinformatics tools for complex data analysis and pattern recognition. The vast amount of spectral data generated by SELDI technology demands implementation of advanced data management and analysis strategies. Another concern remains whether the discovered biomarkers and the derived multivariate models are truly associated with the disease process. Recent reports examined and highlighted the importance of such issues (Baggerly *et al.* 2004; Diamandis 2004a, b). To address these issues, biomarker discovery in the present study was done in two independent ways to cross-validate discoveries: descriptive statistics and the machine-learning algorithms (RSVM scheme). RSVM distinguishes itself from another SVM-based algorithm, the SVM-RFE method (Guyon *et al.* 2002), in three aspects: (1) difference in the scheme of cross-validation, (2) difference in criteria for ranking contributions of a feature to the detection function, and (3) difference in final important marker list is based on frequency-based selection. These three features of RSVM enhance the accuracy of final marker selection, leading to the identification of biomarkers that, in combination, achieved both higher sensitivity and specificity in detecting metal-exposed workers from the non-exposed controls.

Because of the multi-mechanisms of metal toxicology, it is very likely that a combination of several markers will be necessary to effectively detect and diagnose early metal poisoning. Taking advantage of the recent development in SELDI and of the ProteinChip technology, we were able to simultaneously analyze the comprehensive protein profiles in the serum samples from mixed metal-exposed workers and controls. The use of RSVM, which screens potential marker directly on raw spectrum data, allows evaluation of each mass peak according to its cumulative contribution toward the maximal separation of metal mixture exposure from non-exposure. This led to the identification of the five discriminatory biomarkers, with results confirmed by descriptive

statistical analysis. The ROC analysis showed that the AUCs for single selected biomarker were generally moderate, suggesting that the power of single protein peak to discriminate metal exposure from non-exposure was relatively low. However, when multivariate biomarkers were combined to form a single proteomic signature, an increased AUC, along with higher sensitivity and specificity was achieved. The approach of using multiple biomarkers to differentiate disease from non-disease has been used previously in other diseases. For instance, a combination of three serum proteomic markers has been found to reach higher sensitivity and specificity to detect ovarian cancer (Zhang *et al*. 2004). Similarly, a proteomic signature comprising of three serum proteins was reported to be able to detect prostate cancer (Li et al. 2002).

The sensitivity and specificity of the proteomic signature found in this study were somewhat lower than those protein markers detected in cancer subjects (Li *et al*. 2002; Zhang *et al*. 2004), not surprising since we studied asymptomatic, exposed individuals and not patients. Additionally, As and Pb exist in the biological fluids in almost all human beings. The protein signature in the present study may be able to discriminate higher vs. low level (i.e. background) of metal exposure.

It should be pointed out that current knowledge about proteomic biomarkers is just in its infancy. Special cautions should be taken when interpreting the results from the present study. Firstly, the identification of the proteomic signature in this study is not necessary for making a clinical diagnosis. Knowing the identities of the protein biomarkers are, however, more essential from a discovery perspective. Further studies are required to identify the nature of these markers individually and in combination, and to understand the biological role of these proteins in the pathogenesis of metal mixture exposure. Secondly, the classification power of RSVM in protein biomarker screening needs to be reproduced in future independent studies of larger population. Thirdly, the present study focused only on adult males with mixed metal exposure at higher levels, the characteristics of proteomic biomarkers in lower exposure subjects, in females and children require to be further defined. Finally, proteomic biomarkers, like any other molecular biomarkers, can be considered as biological indicators of environmental exposure. Thus, the relationship between proteomic biomarkers and exposure can only be established on the basis of well-defined exposure information.

In summary, we have showed that proteomic technology in combination with bioinformatics tools is a promising novel approach for the discovery of new and better biomarkers for metal mixture exposure. Using multiple serum proteomic markers, we could achieve higher sensitivity and specificity than single marker for detection of higher metal mixture exposure.

## Acknowledgements

The authors thank the participants for their cooperation. We are grateful to Drs. Wei Zhou, Zhaoxi Wang, Geoffrey Liu, and Sohee Park for providing insightful comments; Lia Shimada for research support. This study was supported by Guangxi Institute of Occupational Health, China; Fogarty International Institute of the National Institute of Health (NIH) (USA, grant: TW00828) and grant (ES 00002) from NIH (USA).

## References

Baggerly KA, Morris JS, Coombes KR. 2004 Reproducibility of SELDI-TOF protein patterns in serum: comparing datasets from different experiments. *Bioinformatics* **20**, 777–785.

Bae DS, Gennings C, Carter WH Jr, Yang RSH, Campain JA. 2001 Toxicological interactions among arsenic, cadmium, chromium, and lead in human keratinocytes. *Toxicol Sci* **63**, 132–142.

Binks K, Doll R, Gillies M, Holroyd C, Jones SR, McGeoghegan D, Scott L, Wakeford R, Walker P. 2005 Mortality experience of male workers at a UK tin smelter. *Occup Med* **5**, 215–226.

Chen W, Chen J. 2002 Nested case–control study of lung cancer in four Chinese tin mines. *Occup Environ Med* **59**, 113–118.

China National Criteria Committee (CNCC). 1995 Analytical Methods of lead in the Air of Work Places by Graphite Furnace Atomic Absorption Spectrophotometry. *GB/T16008-1995*.

China National Criteria Committee (CNCC). 1996 Analytical Methods of Arsenic in Urine with Spectrophotometer. *WS/T28-1996*.

China National Criteria Committee (CNCC). 1997 Hygienic Criteria of Arsenic and its Compounds in the Air of Work Places. *GB17055-1997*.

China National Criteria Committee (CNCC). 2002 Diagnostic Criteria for Occupational Chronic Lead Poisoning. *GBZ237-2002*.

De Rosa CT, Johnson BL, Fay M, Hansen H, Mumtaz MM. 1996 Public health implications of hazardous waste sites: findings, assessment, and research. *Food Chem Toxicol* **34**, 1131–1138.

Diamandis EP. 2004a Analysis of serum proteomic patterns for early cancer diagnosis: drawing attention to potential problems. *J Natl Cancer Inst* **96**, 253–256.

Diamandis EP. 2004b Mass spectrometry as a diagnostic and a cancer biomarker discovery tool: opportunities and potential limitations. *Mol Cell Proteomics* **3**, 367–378.

Elsenhans B, Schmolke K, Kolb K, Stokes J, Forth W. 1987 Metal–metal interactions among dietary toxic and essential trace elements in the rat. *Ecotoxicol Environ Saf* **14**, 275–287.

Eng J. ROC analysis: web-based calculator for ROC curves. http://www.rad.jhmi.edu/roc.

Fay M, Mumtaz MM. 1996 Development of a priority list of chemical mixtures occurring at 118 hazardous waste sites using the Hazdat database. *Food Chem Toxicol* **34**, 1163–1165.

Gebel T. 2000 Confounding variables in the environmental toxicology of arsenic. *Toxicology* **144**, 155–162.

Gidlow DA. 2004 Lead toxicity. *Occup Med* **54**, 76–81.

Goyer RA. 1996 Toxic effects of metals. In: Klaassen CD (Ed.), *Casarett & Doull's Toxicology: The Basic Science of Poisons*. McGraw-Hill; New York: pp. 691–736.

Hanley JA, McNeil BJ. 1982 The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Diagnostic Radiol* **143**, 29–36.

Hughes MF. 2002 Arsenic toxicity and potential mechanisms of action. *Toxicol Lett* **133**, 1–16.

International Agency for Research on Cancer (IARC). 2003 *Lists of IARC Evaluations*. Available: http://www.iarc.fr.

Koller K, Brown T, Spurgeon A, Levy L. 2004 Recent developments in low-level lead exposure and intellectual impairment in Children. *Environ Health Perspect* **112**, 987–994.

Li J, Zhang Z, Rosenzweig J, Young Y, Chan DW. 2002 Proteomics and bioinformatics approaches for identification of serum biomarkers to detect breast cancer. *Clin Chem* **48**, 1296–1304.

Lilis R, Valciukas JA, Weber JP, Malkin J. 1985 Effects of low-level lead and arsenic exposure on copper smelter workers. *Arch Environ Health* **40**, 38–47.

Mazur IA, Pazynich VM, Mandrichenko BE, Chinchevich VI. 1983 Spectrophotometric determination of arsine in the atmosphere. *Gig Sanit* **5**, 52–53 (in Russian).

Ministry of Health, China. 2002 Occupational Limit for Hazardous Agents in the Workplaces. *GBZ2-2002*.

National Institute for Occupational Safety and Health (NIOSH). 1994 *NIOSH Manual of Analytical Methods (NMAM)*. Fourth edition, 94–113.

National Institute for Occupational Safety and Health (NIOSH). http://www.cdc.gov/niosh/topics/ABLES/pdfs/ables-lead-data.pdf.

National Research Council. 2001 Arsenic in Drinking Water. Update. National Academy Press; Washington, DC.

Nordberg GF, Anderson O. 1981 Metal interactions in carcinogenesis: enhancement, inhibition. *Environ Health Perspect* **40**, 65–81.

Petricoin EF, Ardekani AM, Hitt BA, Levine PJ, Fusaro VA, Steinberg SM, *et al*. 2002 Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* **359**, 572–577.

Röcken C, Ebert MPA, Roessner A. 2004 Proteomics in pathology, research and practice. *Pathol Res Prac* **200**, 69–82.

Sakai T. 2000 Biomarkers of lead exposure. *Ind Health* **38**, 127–142.

Tsai SM, Wang TN, Ko YC. 1999 Mortality for certain diseases in area with high levels of arsenic in drinking water. *Arch Environ Health* **54**, 186–193.

Vlahou A, Schellhammer PF, Mendrinos S, Patel K, Kondylis FI, Gong L, Nasim S, Wright Jr. GL. 2001 Development of a novel proteomic approach for the detection of transitional cell carcinoma of the bladder in urine. *Am J Pathol* **158**, 1491–1502.

World Health Organization (WHO). 1995 *EHC/IPCS Document: Inorganic Lead (EHC 165)*.

World Health Organization (WHO). 2001 *EHC/IPCS Document: Arsenic and Arsenic Compounds (EHC 224)*.

World Health Organization (WHO). 2003 United Nations Synthesis Report on Arsenic in Drinking Water. *Chapter 2: Environmental Health and Human Exposure Assessment*.

Xu XQ, Leow CK, Lu X, Zhang X, Liu JS, Wong WH, Asperger A, Deininger S, Leung HCE. 2004 Molecular classification of liver cirrhosis in a rat model by proteomics and bioinformatics. *Proteomics* **4**, 3235–3245.

Xiao Z, Prieto D, Conrads TP, Veenstra TD, Issaq HJ. 2005 Proteomic patterns: their potential for disease diagnosis. *Mol Cell Endocrinol* **230**, 95–106.

Zhang Z, Bast RC, Yu Y, Li J, Sokoll LJ, Rai AJ, Rosenzweig JM, Cameron B, Wang YY, Meng XY, Berchuck A, Von Hoaften-Day C, Hacken NF, de Bruyn HW, van der Zee AG, Jacobs IJ, Fung ET, Chan DW. 2004 Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer. *Cancer Res* **64**, 5882–5890.

Zhukov TA, Johanson RA, Cantor AB, Clark RA, Tockman MA. 2003 Discovery of distinct protein profiles specific for lung tumors and pre-malignant lung lesions by SELDI mass spectrometry. *Lung Cancer* **40**, 267–279.