



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Application of proteomics methods for pathogen discovery[☆]

Yiming Ye^{a,1}, Eng-Chun Mar^b, Suxiang Tong^{b,*}, Scott Sammons^a, Sunan Fang^a,
Larry J. Anderson^b, Dongxia Wang^{a,*,1}

^a Biotechnology Core Facility Branch, Division of Scientific Resources, NCPDCID, Atlanta, GA, United States

^b Division of Viral Diseases, NCIRD, Centers for Disease Control and Prevention (CDC), Atlanta, GA, United States

A B S T R A C T

Article history:

Received 23 January 2009

Received in revised form 26 August 2009

Accepted 7 September 2009

Available online 12 September 2009

Keywords:

Detection

Pathogen

Proteomics

Mass spectrometry

Proteomics have been used widely to study proteins in complex materials such as cells, body fluids, tissues, and organisms. Application of advance proteomic techniques for the characterization of disease-specific proteins may provide information for the detection of potential infectious agents. In this report, two proteomics techniques, a two-dimensional differential gel electrophoresis (2D-DIGE) and a one-dimensional gel electrophoresis and one-dimensional liquid chromatography coupled with mass spectrometry (GeLC-MS/MS), were applied for investigating viral proteins from cultured cells inoculated with a clinical sample. The 2D-DIGE method identified five viral proteins of *vaccinia virus* that are only present in infected cells, these results are in agreement with findings determined by genome based methods. The GeLC-MS/MS method identified eight *vaccinia virus* proteins out of 428 proteins detected in the sample. These results demonstrate that proteomic techniques can be used effectively for the detection of infectious agents. Given that the methods are capable of applying to proteins without a prior knowledge of the pathogen present, proteomics has a potential of being developed as a molecular tool for pathogen discovery, and disease diagnosis of emerging infectious diseases and for bioterrorism defense.

Published by Elsevier B.V.

1. Introduction

Proteomics are a powerful tool for the study of proteins in complex mixtures including cells, body fluids, tissues, organisms and other complicated samples in the post-genomic era (Aebersold and Mann, 2003). Advances in mass spectrometry combined with multi-dimensional orthogonal separation techniques and genomic databases allow qualitative and quantitative determination of protein expression, post-translational modification, and protein–protein interaction rapidly and with high throughput.

Qualitative protein profiling of the complex and dynamic entities that constitute a proteome can be accomplished by multiple-dimensional gel electrophoresis, high performance liquid chromatography (HPLC), and combinations of these techniques coupled with mass spectrometry (MS). In a gel-based approach, complex protein mixtures can be resolved effectively according to their isoelectric points and molecular weights by two-dimensional polyacrylamide gel electrophoresis (2D-GE) (O'Farrell, 1975). Sep-

arated protein spots are then excised from the gel and subjected to mass spectrometry for identification. While 2D-GE is a separation technique used widely prior to MS analysis, two- or multi-dimensional liquid chromatographic techniques have been developed and used increasingly in analyzing complex protein mixtures. In a shotgun method termed multi-dimensional protein identification technology (MudPIT) (Washburn et al., 2001), protein mixtures are digested first into a pool of peptide mixtures without prior sample fractionation. Next, peptides are fractionated using two-dimensional liquid chromatography (LC), which consists typically of a strong cation exchange chromatography (SCX) as the first dimension and a reverse phase chromatography separation as a second dimension. Thousands of protein components in cell lysates or other complex systems can be characterized directly by this approach. In addition, the gel and LC based technologies can be combined such that one-dimensional PAGE gel electrophoresis is used to separate proteins, and nano-capillary LC-MS/MS analysis is used to analyze the peptides generated from these proteins after in-gel digestion (GeLC-MS/MS) (Wilm et al., 1996; Schirle et al., 2003). This approach is becoming increasingly popular in the field of proteomics research due to its combination of the robust nature of SDS-PAGE and the resolving power of LC-MS/MS.

Another challenge in proteomics has been to identify the relative abundance of proteins between two or more samples, and a number of techniques have been developed for this purpose (Mann

[☆] The findings and conclusions in this report are those of the author(s) and do not necessarily represent the views of the funding agency.

* Corresponding authors. Tel.: +1 770 488 0446; fax: +1 404 639 1331.

E-mail address: dov2@cdc.gov (D. Wang).

¹ These authors contributed equally.

and Aebersold, 2003). 2D-GE was a classic method used extensively in the early date of proteomics. Relative abundance of proteins between two or more samples, or unique proteins present only in one sample, can be determined by comparing directly the intensity of distinct gel spots between two or more gels. With 2D-GE, however, gel-to-gel variation, location of spots and low dynamic range of standard colorimetric staining techniques, make it difficult to identify confidently differences in protein composition between samples. These limitations can be overcome by a two-dimensional differential gel electrophoresis technique (2D-DIGE) (Ünlü et al., 1997) developed recently. In this method, two sets of protein extracts are labeled with two distinct and spectrally resolvable fluorescent tags and the mixture of labeled proteins is then applied onto a single 2D-GE gel separation. Differentially expressed or sample-specific proteins are detected and analyzed based on the distinctive colored spots. Since both samples are run on the same gel, gel-to-gel variability does not confound the results. An added advantage is that the sensitivity as well as linearity in the amounts of protein detected can be increased by fluorescence labels. Besides gel based methods, other quantitative proteomics techniques used commonly include stable isotopic chemical labeling of peptides or proteins such as isotope-coded affinity tags (ICAT) (Gygi et al., 1999) and isobaric tags for relative and absolute quantitation (iTRAQ) (Ross et al., 2004), metabolic labeling such as stable-isotope labeling of amino acids in cell culture (SILAC) (Ong et al., 2002), enzyme-catalyzed ^{18}O incorporation (Yao et al., 2001), and label-free spectra comparison (Chong et al., 2001).

Recent development in proteomic techniques has provided the means for investigating various aspects of infectious diseases, including characterization of the proteome of infective agents, assessment of global cellular response toward infection, identification of potential anti-infection drug targets, and development of new diagnostic methods. A number of groups have applied proteomics to the study of virus infections (Maxwell and Frappier, 2007). For instance, the protein components of intracellular mature virions (IMV) from *vaccinia virus* have been investigated using various proteomics approaches and up to 80 viral proteins were identified (Chung et al., 2006; Yoder et al., 2006; Resch et al., 2007). Shotgun and ICAT proteomic methods were employed for identifying 28 white spot syndrome virus (WSSV) encoded proteins and 12 infection responsive cellular proteins (Wu et al., 2007). Recently, a proteomic fingerprinting method using surface-enhanced laser desorption/ionization time-of-flight (SELDI-TOF) mass spectrometry was developed for potential application to early diagnosis of severe acute respiratory syndrome (SARS) (Kang et al., 2005).

Key to the study of human infectious diseases is identification of the causative agent. In this report, proteomic techniques for the detection and characterization of infectious agents were applied to complement methods that detect antibody responses, antigens, or nucleic acids associated with the pathogen. As a proof of concept, a cell culture was inoculated with an unknown infectious patient's sample (Anonymous, 2007) and was examined using two proteomics methods, 2D-DIGE and GeLC-MS/MS. The comparative 2D-DIGE method led to the direct identification of five *vaccinia virus*-encoded proteins in infected cells, but not in uninfected control cells. This is consistent with the finding of the *vaccinia virus* by a deoxyribonuclease sequence-independent single primer amplification (DNase-SISPA) sequencing method (Anonymous, 2007). In addition, high sequence coverage of the viral protein E3 provides information to determine the specific *vaccinia virus* strain. These results suggest that proteomic techniques could be used as a potential powerful molecular tool for pathogen discovery and diagnostic, particularly important with the emergence of infectious diseases and the bioterrorism threat.

2. Materials and methods

2.1. Virus and cells

An unidentified viral culture isolate was inoculated with a vulvar swab specimen from a labial lesion of the patient (Anonymous, 2007) and was then sent to CDC for investigation. The human lung fibroblastic cells (HLF), obtained from CDC, were used for further inoculation. The cells (passages 16–18) were subcultured in Eagle's minimum essential medium (EMEM) (MediaTech, Cellgro, VA, USA) supplemented with 10% fetal calf serum (FCS) (Hyclone, Logan, UT, USA), 100 U of penicillin, and 100 μg of streptomycin per milliliter.

Twenty microliters of this isolate in 2 mL of EMEM was added to the HLF cells growing exponentially in a 25 cm^2 -flask. After adsorption for 1 h, the inoculum was removed and replenished with 10 mL of EMEM containing 5% heat-inactivated FCS with the antibiotics. Cells were harvested 36 h post-infection when approximately 60% of cells showed CPE. After washing with cold PBS (0.01 M sodium phosphate, 0.15 M NaCl, pH7.4), cells were scraped into PBS and centrifuged at 2000 rpm for 5 min at 4 °C. The cell pellet was subject to protein extraction. Sham-infected cells were also obtained as control.

2.2. Preparation of cell extracts

Sham- or virus-infected cell pellet was disrupted in RIPA buffer (50 mM Tris, pH 7.4, 150 mM NaCl, 0.5% Nonidet-P 40, 0.1% SDS, 0.5% sodium deoxycholate, 5 mM EDTA, and 1 mM PMSF (phenylmethylsulfonyl fluoride) at a concentration of 1×10^7 cells/mL, and placed on ice for 20 min. The resultant mixture was vortexed briefly and centrifuged at 27,216 $\times g$ for 10 min at 4 °C. The supernatant was collected and stored at -80°C until use.

2.3. Two-dimensional fluorescence differential gel electrophoresis (2D-DIGE)

Two sets of extracted samples (infected and uninfected) were processed with 2D-DIGE analysis. The protein concentrations were determined using 2D-Quant Kit from GE healthcare. 100 μL of each cell lysate (80 μg) was prepared first with a 2-D Clean Up Kit (GE healthcare, Piscataway, NJ, USA) to remove interfering components. The proteins were then labeled according to the manufacturer's protocol (GE healthcare, Piscataway, NJ, USA). In brief, infected and control proteins were reconstituted in a labeling buffer, and labeled with 400 pmol of Cy3 (infected protein) and Cy5 (control protein) fluorescence dyes, on ice for 30 min in the dark. The labeling reaction was terminated with 10 mM lysine for 10 min. The labeled samples were combined and mixed in a rehydration buffer (7 M urea, 2 M thiourea, 2% CHAPS, 50 mM DTT, 1% Pharylyte) and run in a single gel. The samples were rehydrated for 12 h and were focused on an 11-cm immobilized pH gradient strip (pH 3–11 NL, GE-healthcare) on a PROTEAN IEF cell (Bio-Rad, Hercules, CA, USA). Isoelectric focusing was performed under the condition: 500 V constant for 500 Vh; linear gradient to 1000 V for 800 Vh; gradient to 6000 V for 7000 Vh; gradient to 6000 V for 3700 Vh. After the first-dimensional isoelectric focusing, IPG strips containing proteins were treated for 15 min with gentle shaking in the equilibrium buffer (6 M urea, 30% glycerol, 2% SDS, 50 mM Tris, pH 8.8) containing 2.5% tributyl phosphine (TBP) to reduce the proteins, followed by protein alkylation with iodoacetamide (3%) for 15 min. The second-dimension SDS-PAGE was run on an 8–16% linear gradient criterion Tris-HCl gel (Bio-Rad) at 110 V for 2 h. For a conventional two-dimensional gel electrophoresis experiment, the protein lysates of infected and control cells were run separately on two individual SDS-PAGE gels after the first isoelectric focusing step under the condition similar to the 2D-DIGE method.

The gel images were obtained using Typhoon 9400 imager (GE-healthcare) at appropriate wavelengths for Cy3 and Cy5 dyes and analyzed with the Master Imager software (GE-healthcare). The gels were then visualized by colloidal Coomassie staining or silver-staining. The protein spots of interest were excised for further analysis.

2.4. Sample preparation for one-dimensional gel electrophoresis and one-dimensional liquid chromatography coupled with mass spectrometry (GelC-MS/MS) analysis

After cleaning with the above mentioned 2-D Clean Up kit, the protein extracts of infected cells were precipitated, solubilized in lysis buffer (1 M Tris-HCl, 8 M urea and 4% CHAPS, pH 8.5) and run on an 8–16% linear gradient criterion Tris-HCl gel (Bio-rad) at 110 V for 1 h. After colloidal Coomassie staining of the gels, the protein lanes between 10 and 50 kDa in size were sliced into 25 pieces in equal size for further analysis.

2.5. Peptide synthesis

Peptides, VAVC.ACAM2000 peptide_{66–96} (WFMTEADKPDADV-MADAIIDDVSR) and VAVC.COP.E3L peptide_{66–96} (WFMTEADKPDADAMADVIIDDVSR) were prepared with an Fmoc chemistry using solid phase peptide synthesis method. Peptides were purified by reversed phase (C-18) high performance liquid chromatography using a gradient of water-acetonitrile (0.1% trifluoroacetic acid). The correct structures of the peptides were confirmed by mass spectrometry.

2.6. Mass spectrometry analysis

Sample in-gel digestion and micro-purification were carried out on a ZipPlate micro-SPE according to the manufacturer's In-Gel Digestion Protocol (Millipore, Billerica, MA, USA) with minor modifications. Briefly, gel spots or bands were destained in buffer 1 (25 mM ammonium bicarbonate and 5% acetonitrile) for 30 min and then in buffer 2 (25 mM ammonium bicarbonate and 50% acetonitrile) for an additional 30 min. After reduction with 10 mM DTT and alkylating with 55 mM IAA, the gel pieces were incubated in 30 μ L of the trypsin solution (5 ng/ μ L trypsin in 25 mM ammonium bicarbonate) and incubated overnight at 37 °C. During digestion with other endoproteases, the reactions were conducted at different solutions (30 μ L of 5 ng/ μ L Lys-C, Glu-C or Arg-C in 100 mM ammonium bicarbonate) and incubated at 37 °C overnight. The in-gel digested peptides were extracted, micro-purified, vacuum-dried and resuspended in 1% formic acid and 2% acetonitrile for mass spectral analysis. Nano-capillary LC-MS/MS analysis was performed in a Micromass Q-ToF Ultima mass spectrometer equipped with a nanospray ion source and coupled with a nanoAcquity ultraperformance liquid chromatography system (UPLC) (Waters, Milford, MA, USA). The protein digest (2 μ L) was loaded onto an "in-house" packed reverse phase C18 capillary column (~15 cm, 75 μ m inner diameter) and separated using a linear gradient of 15–45% of buffer B (acetonitrile, 0.1% formic acid) in buffer A (water, 0.1% formic acid) in 60 min (2D gel samples), or 80 min (1D gel samples) at a flow rate of 0.5 μ L/min.

All mass spectra were obtained in the positive-ion mode. An electric potential of 3.5 kV was applied to the emitter in the ion source. The acquisition of data was performed on a MassLynx data system (version 4.0) using a data-dependent mode where the four highest intensity precursors in an MS1 survey scan were isolated for collision-induced dissociation. The resulting MS/MS data was searched for protein candidates with automated database searching against NCBI nr database using MASCOT Daemon software (Matrix Sciences, London, UK). The mass tolerance of both precursor

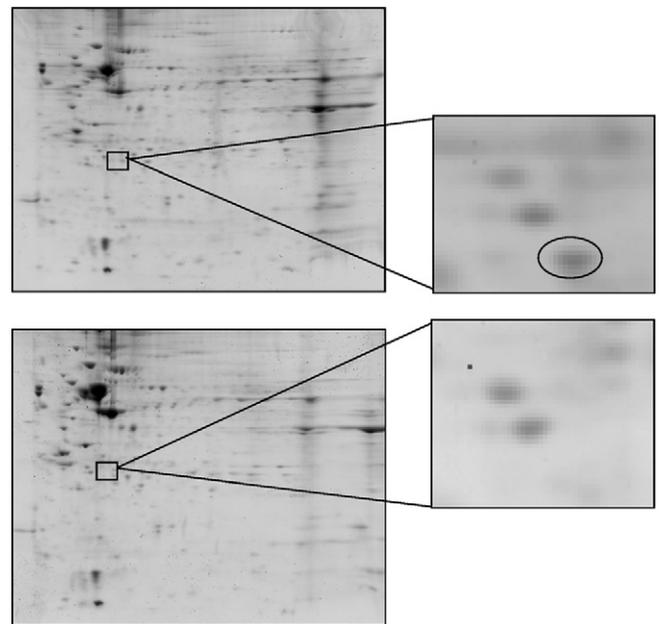


Fig. 1. Two-dimensional gel electrophoresis of proteins extracted from vaccinia virus-infected (upper) and uninfected (bottom) cells. Amplified regions show different protein patterns and the spot corresponding to one distinctive protein is circled.

ions and fragment ions was set to ± 0.4 Da. Carboxyamidomethyl cysteine and oxidized methionine were set as variable modifications. All viral peptides identified during MASCOT searches were confirmed by manual inspection.

2.7. Sequence analysis

CLUSTALW 1.83 (Thompson et al., 1994) was used to align the amino acid sequences of the Vaccinia Copenhagen E3 protein and homologues from 67 poxviruses. Sequences that were identical were condensed into one row for readability. The alignment was formatted using the BioEdit sequence alignment editor (Hall, 1999).

3. Results

3.1. Identification of viral proteins by 2D-DIGE method

A virus isolate obtained from an infected patient was inoculated in human lung fibroblast (HLF) cells. The protein mixtures were extracted from unknown virus-infected cells and mock-infected control cells. In the initial experiment, two sample sets were first separated by conventional two-dimensional gel electrophoresis techniques (2-DE). Only one unique protein spot was visualized through comparison of the 2-DE of infected to uninfected cell lysate (Fig. 1). The protein corresponding to the spot was identified as vaccinia virus-encoded protein E3 after in-gel trypsin digestion followed by LC/MSMS analysis of extracted peptides, consistent with the results obtained by a DNase-SISPA sequencing method and an epidemiologic investigation that the patient was infected by a smallpox vaccine-strain vaccinia virus (Anonymous, 2007).

The samples were then analyzed by a 2D-DIGE method, a proteomic technique developed recently. The infected cell lysate was labeled with Cy5, whereas the lysate from uninfected control cells was labeled with Cy3. The mixture of labeled sample pair was resolved in a single 2D gel. The fully resolved spots are shown in Fig. 2, which was generated by overlapping two fluorescent images corresponding to the protein profiles of the infected and control samples. The yellow spots in the image represent the overlapping

Table 1
Proteins identified from selected DIGE spots by LC–MS/MS.

Spot no.	Protein name	Species	ORF name	Classification of viral gene	M.W. (kDa)	MASCOT score ^a	Matched peptides	Sequence coverage (%)
1	Protein E3	<i>Vaccinia virus</i>	E3L	Early	21.4	1018	18	77
2	Protein B2	<i>Vaccinia virus</i>	B2R	Early-late	24.6	105	6	31
3	Serum albumin precursor	Human			69.3	1089	29	53
4	Serum albumin precursor	Human			69.3	695	22	39
5	Glutaredoxin-1	<i>Vaccinia virus</i>	O2L	Late	12.3	63	3	25
6	14 kDa fusion protein	<i>Vaccinia virus</i>	A27L	Late	12.6	34	6	29
7	Cytochrome c oxidase subunit VIb isoform 1	Human			10.1	96	2	23
8	Glutaredoxin-2	<i>Vaccinia virus</i>	G4L	Late	13.9	68	3	23

^a Probability based mowse score that indicates the quality of the MS/MS peptide fragment ion matches.

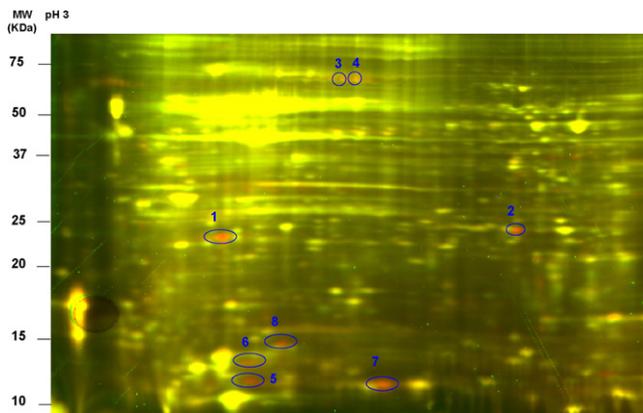


Fig. 2. Two-dimensional differential gel electrophoresis (2D-DIGE) of protein mixtures extracted from infected and control cells. The infected sample is labeled with Cy5 (displayed as red color) and the uninfected sample with Cy3 (displayed as green color).

of red (infected) and green (control), indicating an equivalent or similar abundance of proteins in both uninfected and infected cells. In contrast, the red-colored spots suggest the protein was specific to the infected cell lysate and spots with intermediate shades of yellow and red probably corresponds to differential levels of host protein expression in response to infection. With the aim for detecting the infectious pathogen(s), this study focused mainly on the proteins identified solely in infected cells with the expectation that they might be virus-encoded proteins.

Eight red spots were selected for in-gel trypsin digestion and mass spectrometry analysis, as described in the methods section. The tryptic digests of each spot were analyzed by the LC coupled tandem mass spectrometry (MS/MS) using a Q-TOF MS instrument. Fig. 3 shows a typical tandem mass spectrum of the peptides digested from the protein of spot 1. As shown in Fig. 3, the high-quality mass spectrum containing a complete series of y-type fragment ions allows the unambiguous identification of the peptide, NIGIEGATAAQLTR, derived from trypsin digestion of Protein E3 of the *vaccinia virus*. Using the MASCOT algorithm against the NCBI nr database, five of the eight proteins from the selected gel spots were identified as viral proteins of smallpox *vaccinia virus* (Table 1). Any matched peptides with low probability scores were double-checked by manual inspection to avoid false positive results.

In order to characterize further the isolate, sequence analysis or alignment on all identified viral proteins was conducted. In the search of the NCBI nr database, one set of peptides could match to one or more proteins, including isomers that share the sequences of identified peptides. The mass spectra data sets of the spots 4, 5, and 8 resulted in three viral proteins (glutaredoxin-1, -2 and 14 kDa fusion protein) that could be from either *vaccinia virus* or other poxvirus species because the sequences of these virus-encoded proteins are either highly conserved or the virus- or strain-specific regions were not included in the identified peptides (data not shown). In contrast, the data of spots 1 and 2 were found to be *vaccinia virus*-encoded protein E3 and B2. Fig. 4 shows the alignment of protein E3s among all poxvirus and strains where identical sequences were grouped into one row. As indicated by underlined amino acids, the peptides detected in the

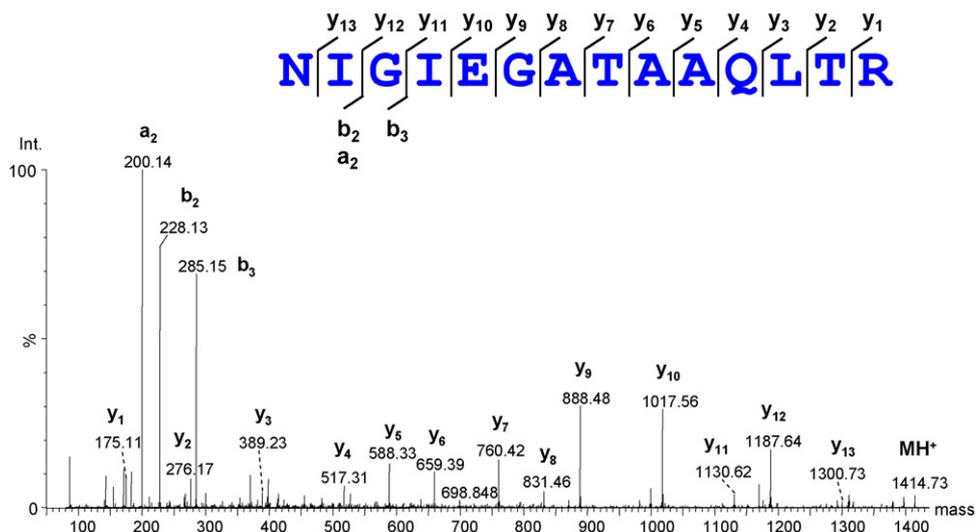
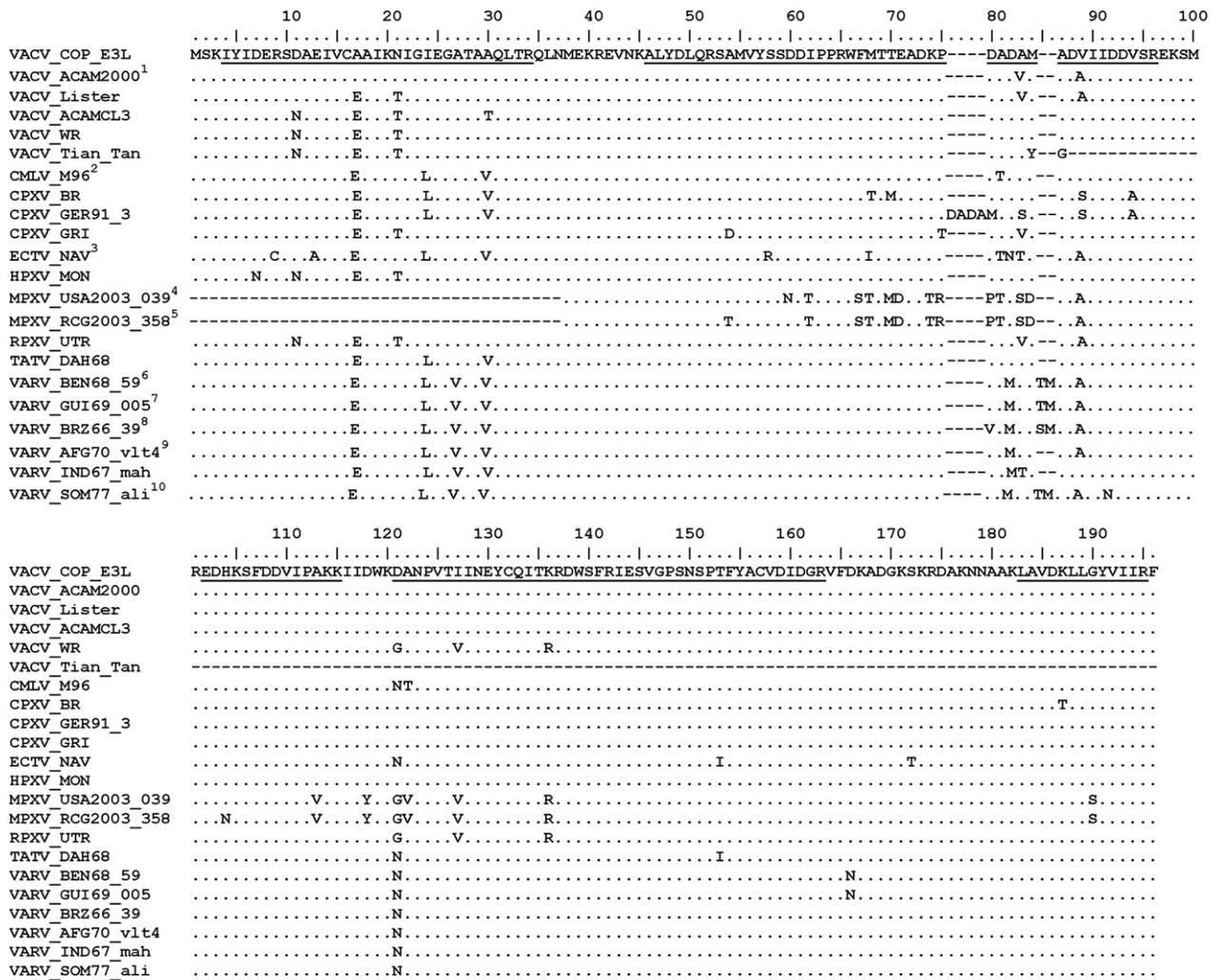


Fig. 3. A typical MS/MS spectrum of a peptide detected from the digested protein in spot 1 as shown in the Fig. 2. y-type ions represent the C-terminal fragments of the peptide whereas b- and a-type ions designate N-terminal fragments (Roepstorff and Fohlman, 1984).



1. include *vaccinia virus* VACV_ACAM2000 and VACV_MVA
2. include *cowpox virus* CMLV_M96 and CMLV_CMS
3. include *ectromelia virus* ECTV_NAV and ECTV_MOS
4. include *monkeypox virus* MPXV_USA2003_039, MPXV_1961_WR, MPXV_LIB1970_184
5. include *monkeypox virus* MPXV_RCG2003_358, MPXV_ZAI96, MPXV_ZAI79_005
6. include *variola virus* VARV_BEN68_59, VARV_BOT72_143, VARV_BOT73_225, VARV_BSH74_nur, VARV_BSH74_shz, VARV_BSH74_sol, VARV_BSH75_banu, VARV_CHN48_horn, VARV_CNG70_227, VARV_CNG70_46, VARV_ETH72_16, VARV_ETH72_17, VARV_GER58_hdlg, VARV_KOR47_lee, VARV_NIG69_001, VARV_IND53_mad, VARV_SAF66_102, VARV_SAF65_103, VARV_SUD47_jub, VARV_SUD47_rum, VARV_SUM70_222, VARV_SUM70_228, VARV_TAN65_kem
7. include *variola virus* VARV_GUI69_005 and VARV_SLN68_258
8. include *variola virus* VARV_BRZ66_39 and VARV_BRZ66_gar
9. include *variola virus* VARV_AFG70_vlt4, VARV_IND53_ndel, VARV_IND64_vel4, VARV_IND64_vel5, VARV_IRN72_tbrz, VARV_JAP46_yam, VARV_JAP51_hrpr, VARV_JAP51_stwl, VARV_JAP51_ytm, VARV_KUW67_1629, VARV_NEP73_175, VARV_PAK69_lah, VARV_SYR72_119 and VARV_YUG72_164
10. include *variola virus* VARV_SOM77_ali, VARV_SOM77_1252 and VARV_SOM77_1605

Fig. 4. Sequence alignment of protein E3 in 67 poxvirus species and strains using the sequence of the protein in the Copenhagen strain of *vaccinia virus* (VACV.COP.E3L) as a reference. The identified regions are underlined. Identical sequences are condensed into a single row for readability. The sequences included are: *vaccinia viruses* VACV_COP (NC.001559), VACV_ACAM2000 (AY313847), VACV_MVA (U94848), VACV_Lister (AY678276), VACV_ACAMCL3 (AY313848), VACV_WR (AY243312), and VACV_Tian_Tan (AF095689); *Camelpox viruses* CMLV_M96 (AF438165) and CMLV_CMS (AY009089); *Cowpox viruses* CPXV_BR (AF482758), CPXV_GER91_3 (DQ437593), and CPXV_GRI (X94355); *Ectromelia viruses* ECTV_NAV (Naval) and ECTV_MOS (NC.004105); *horsepox virus* HPXV_MON (DQ792504); *Monkeypox viruses* MPXV_USA2003_039 (DQ011157), MPXV_1961_WR (AY603973), MPXV_LIB1970_184 (DQ011156), MPXV_RCG2003_358 (DQ011154), MPXV_ZAI96 (NC.003310), and MPXV_ZAI79_005 (DQ011155); *Rabbitpox virus* RPXV_UTR (AY484669); *Taterapox virus* TATV_DAH68 (DQ437594); *Variola viruses* VARV_BEN68_59 (DQ441416), VARV_BOT72_143 (DQ441417), VARV_BOT73_225 (DQ441418), VARV_BSH74_nur (DQ441420), VARV_BSH74_shz (DQ441421), VARV_BSH74_sol (DQ441422), VARV_BSH75_banu (DQ437581), VARV_CHN48_horn (DQ437582), VARV_CNG70_227 (DQ441423), VARV_CNG70_46 (DQ437583), VARV_ETH72_16 (DQ441424), VARV_ETH72_17 (DQ441425), VARV_GER58_hdlg (DQ437584), VARV_KOR47_lee (DQ441432), VARV_NIG69_001 (DQ441434), VARV_IND53_mad (DQ441427), VARV_SAF66_102 (DQ441435), VARV_SAF65_103 (DQ441436), VARV_SUD47_jub (DQ441440), VARV_SUD47_rum (DQ441441), VARV_SUM70_222 (DQ437591), VARV_SUM70_228 (DQ441442), VARV_TAN65_kem (DQ441443), VARV_GUI69_005 (DQ441426), VARV_SLN68_258 (DQ441437), VARV_BRZ66_39 (DQ441419), VARV_BRZ66_gar (Y16780), VARV_AFG70_vlt4 (DQ437580), VARV_IND53_ndel (DQ441428), VARV_IND64_vel4 (DQ437585), VARV_IND64_vel5 (DQ437586), VARV_IRN72_tbrz (DQ437587), VARV_JAP46_yam (DQ441429), VARV_JAP51_hrpr (DQ441430), VARV_JAP51_stwl (DQ441431), VARV_KUW67_1629 (DQ441433), VARV_NEP73_175 (DQ437588), VARV_PAK69_lah (DQ437589), VARV_SYR72_119 (DQ437592) and VARV_YUG72_164 (DQ441448), VARV_SOM77_ali (DQ437590), VARV_SOM77_1252 (DQ441438), VARV_SOM77_1605 (DQ441439), VARV_IND67_mah (NC.001611).

Table 2
Vaccinia virus encoded viral proteins identified by GeLC-MS/MS method.^a

No.	Protein name	ORF	Classification of viral gene	M.W. (kDa)	MASCOT score	Matched peptides	Sequence coverage (%)
1	36 kDa late protein I1	I1L	Late	35.8	164	5	16
2	Protein I3	I3L	Early-late	30.0	33	2	7
3	Structural protein VP8 precursor	L4R	Late	28.4	59	2	8
4	Thymidylate kinase	A48R	Immediate-early	23.2	33	1	5
5	Envelope protein (Protein H5)	H5R	Late	22.2	98	3	24
6	Protein E3	E3L	Early	21.5	107	7	41
7	Protein C16/B22	C16L/B22R	Early	21.0	203	6	33
8	Protein C6	C6L	Early	17.4	99	1	7

^a Total of 274 human proteins and 8 viral proteins were identified from 6188 MS/MS spectra.

regions of E3 had almost all of the amino acid variations suggestive of the Copenhagen strain (VAVC.COP.E3L). However, a minor difference in the E3 sequences between the Copenhagen strain and the other two strains in the group of VACV.ACAM2000 was noted during manual inspection. Their protein sequences are almost identical except a swap of the alanine and valine at the position 83 and 89. The peptide at m/z 938.08 was identified as VAVC.COP.E3L peptide_{66–96} (WFMTEADKPDADAMADVIIDDVSR), by MASCOT database searching, that included the amino acid residues at these two positions. However, the sequence of this peptide was further determined to be derived from VAVC.ACAM2000 encoded protein E3 by mass spectrometric analysis of two synthetic peptides (Fig. 5). The tandem mass spectrum of the peptide digested from the gel spot matches the spectrum of the synthetic peptide derived from the protein of VACV.ACAM2000 rather than the one of Copenhagen strain. Therefore, the strain-specific identification of the infectious pathogen can be narrowed down confidently to two vaccinia virus strains, VACV.ACAM2000 (AY313847) and VACV.MVA (U94848), included in the VACV.ACAM2000 group (Fig. 4).

3.2. Improvement of protein sequence coverage through multiple enzyme digestion

In order to obtain high sequence coverage of viral proteins, the effects of the multiple enzymatic digestions on the protein of spot 1 (Fig. 2) from 2D-DIGE experiments were examined. Apart from trypsin digestion, the identical spots, corresponding to the protein E3, obtained from three parallel experiments were subjected to in-gel digestion individually by three other proteases, Lys-C, Glu-C, and Arg-C that cleave proteins at the C-terminal sites of the residues of lysine, glutamic acid and arginine, respectively. It was demonstrated that the combination of four enzyme digestions increased sequence coverage of the protein from 77% to 89% (Fig. 6).

3.3. Identification of viral proteins using the GeLC-MS/MS method

In order to develop an alternative proteomics method for detecting infective virus, a different approach named GeLC-MS/MS was explored in the present study. Proteins extracted from infected cells were first separated by one-dimensional SDS-PAGE gel and the gel lane containing proteins in the range of approximate 10–50 kDa was excised into 25 equally spaced gel slices (Fig. 7). Nanocapillary LC-MS/MS analysis of the peptides generated from in-gel digestion of all gel bands resulted in the identification of a total of 274 proteins. The majority of the identifications are host cellular proteins (data not shown) but eight vaccinia virus viral proteins were detected, indicating that GeLC-MS/MS method is also capable of providing identity information of on a certain virus infection (Table 2).

4. Discussion

The 2D-DIGE method proved to have higher sensitivity, reproducibility and efficiency than the 2-DE comparison. In comparison to the conventional 2D-gel method, the 2D-DIGE technique identified four more virus-encoded proteins, suggesting that the latter approach is more sensitive, reliable, and can be the choice of method for pathogen discovery with higher confidence. Spot 1 on the 2D-DIGE gel was located in a similar position to the viral protein spot seen by the regular 2-DE gel. Both spots were determined to be protein E3, indicating the consistency of the two methods. The reason that E3 was the only viral protein detected in the conventional 2-DE method could be due to its high abundance revealed by the high score and high sequence coverage of this protein (Table 1). Notably, spots 3, 4, and 7 were identified as three human proteins with no trace of viral peptide sequences. These proteins may be induced by the infection and may not be present in sufficient quantities to be detected in the uninfected control sample.

It was noted from the sequence alignment (Fig. 4) that the primary sequences of the protein E3 between the Copenhagen strain and the other two strains in the group of VACV.ACAM2000 are almost identical except a swap of the alanine and valine at the positions 83 and 89. The identification of the tryptic peptide at m/z 938.08 as a Copenhagen strain peptide, VAVC.COP.E3L peptide_{66–96}, by MASCOT database searching could be problematic and needs further confirmation. The swap of the two amino acid residues within a single peptide does not alter its molecular weight and may not cause a significant change on its fragmentation pattern which is used to determine peptide sequence. The tandem mass spectrum of this precursor ion and the computer assisted sequence assignment was shown in Fig. 5A. Although the majority of experimental fragments match to the different types of ion fragments cleaved theoretically from the assigned sequence, three large fragments at m/z 888.49, 1003.47 and 1074.56, could not be assigned to the putative sequence. In addition, no fragments were observed to cover the region (AMADV) containing those two amino acids that are involved in residue swapping. An effective way to validate the sequence determination of a tryptic peptide is to prepare and analyze a synthetic peptide with the same amino acid sequence as the putative one. Synthetic and tryptic peptides should yield a very similar MS fragmentation pattern if they have identical composition and sequence. For this purpose, two synthetic peptides were prepared, with the sequences derived from the positions 66–96 of two E3 proteins. MS analysis of the synthetic peptides indicated that the three unassigned fragments in Fig. 5A appear on the spectrum of the synthetic VAVC.ACAM2000 peptide_{66–96} (Fig. 5B) but not on the other (Fig. 5C) while all other informative fragments of the tryptic peptide were also observed on this spectrum, indicating that the tryptic peptide was digested from VAVC.ACAM2000 encoded protein E3. Therefore, the infectious pathogen was iden-

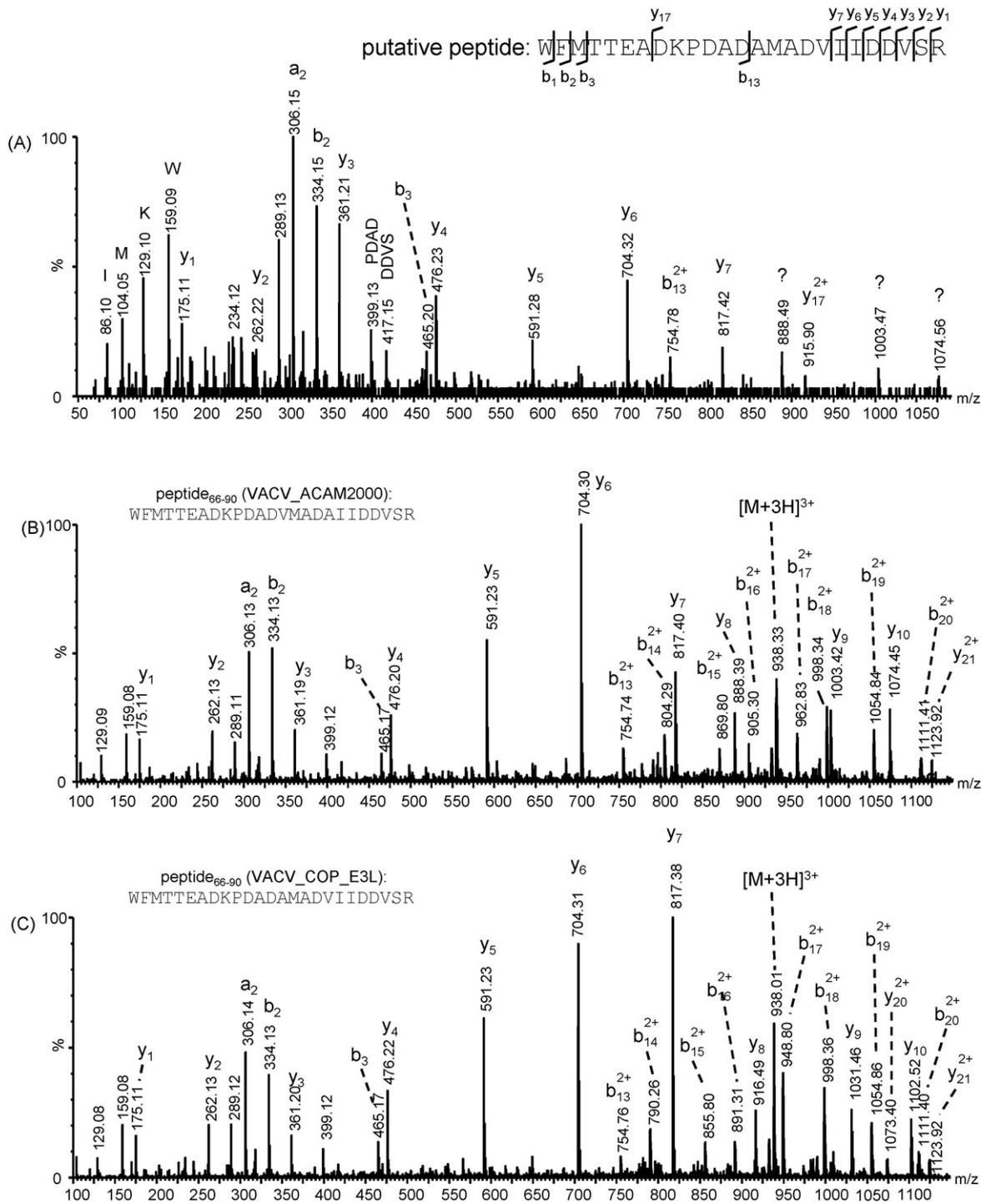


Fig. 5. MS/MS spectra of (a) the tryptic peptide at m/z 938.08, (b) synthetic VAVC.ACAM2000 peptide₆₆₋₉₆, WFMTTEADKPKDADVMADAIIDDVSR, and (c) synthetic VAVC.COP.E3L peptide₆₆₋₉₆, WFMTTEADKPKDADAMADVIIIDDVSR. All fragments were labeled including y-, b-, a-type ions, immonium and internal ions (nomenclature see reference: Roepstorff and Fohlman, 1984).

tified to be either of the two *vaccinia virus* strains included in the VACV.ACAM2000 group (Fig. 4).

Protein E3 is a double-stranded RNA-binding protein that is involved in a mechanism to evade cellular antiviral host defense. Given its considerable abundance in infected cells (Figs. 1 and 2) and its wide variation in amino acids, among various orthogonal poxviruses, it is reasonably expected that protein E3 can serve as a biomarker for the detection of poxvirus infection. Several of the peptides digested from E3 and detected by mass spectrometry resulted in significant MASCOT peptide scores, indi-

ating their high ionization and fragmentation efficiency (data not shown). The combination of these critical criteria and advanced proteomics techniques, such as single- and multiple-reaction monitoring (SRM/MRM) methods (Kuhn et al., 2004), should allow the development of high throughput screening method for direct and sensitive detection of infective virus.

To detect and identify infectious agents and their strains, it is necessary to obtain complete or very high sequence coverage of viral proteins. However, it is often inefficient in normal proteomic procedures to carry out a single enzymatic digestion prior

1 MSKIYIDERSDAEIVCAAIAKNIKIGEGATAAQLTRQLNMEK

 41 REVNKALYDLQRSAMVYSSDDIPRWFMTTEADKPDADAM

 81 ADVIIDDVSREKSMREDHKSFDVVI PAKKI IDWKDANPVT

 121 IINEYCQITKRDFSRIESVSGPSNSPTFYACVDIDGRVFD

 161 KADGKSKRDAKNNAAKLAVDKLLGYVIIRF

Fig. 6. The sequence regions of protein E3 identified through the digestions of multiple enzymes including trypsin (—), Lys-C (···), Arg-C (---) and Glu-C (- - -).

to MS analysis. For example, a single trypsin digestion that cleaved proteins at the lysine and arginine sites usually produces some undetectable tryptic fragments of a small or large size, or low ionizable peptides and therefore only partial sequences of a protein can be determined. This limitation can be overcome by a multiple enzyme digestion strategy that has proved to be an effective way in identification of post-translational modifications on specific proteins (Aebersold and Mann, 2003). In this method, a sample can be divided into two or more fractions and each fraction can be digested with different endoproteases. The multi-protease digestion generates overlapping peptides spanning over much broader regions of target proteins through cleavages at various residue-specific sites. In this study, the effects of the multiple enzymatic digestions on the protein of the spot 1 from 2D-DIGE experiments (Fig. 2) were examined. In comparison to the single trypsin digestion, the combination of four individual digestions with different enzymes leads to more detectable peptides and hence to the increase in the sequence coverage of the target protein (Fig. 6). This should be particularly important for identifying strain-specific virus where the amino acid variation of a protein within the species is small, requiring high sequence coverage for a confident identification of the viral strain.

It is interesting to observe that different sets of viral proteins, except protein E3, were determined by 2D-DIGE and GeLC-MS/MS methods. In addition, GeLC-MS/MS detected more virus encoded proteins than 2D-DIGE did whereas 2D-DIGE method yielded higher probability scores and the number of matched peptides for individual proteins than GeLC-MS/MS technique. These differences can be attributed to the two techniques of intrinsically different protein separation mechanisms and analysis strategies. In the 2D-DIGE method, only a few well-resolved and visualized protein spots were selected for further analysis whereas the other regions of interest are buried in poorly separated spots and could be ignored. In contrast, all proteins in selected gel bands were analyzed in GeLC-MS/MS method and identification depended on chromatographic separation, ionization and fragmentation efficiency of the in-gel generated peptides. Proteins isolated in 2D gel spots may not be detectable in a mixture present in a 1D gel band and vice versa. On the other hand, individual 2D-gel spots usually contain only one or a few purified proteins but more proteins can be present in the 1D-gel bands. Without additional separation steps, fewer peptides derived from gel spots than from gel bands could be expected and this could lead to a better separation and relative higher qualities of tandem mass spectra in a single LC-MS/MS run. Overall, both proteomics methods have their own advantages. 2D-DIGE produces straight forward images for the selection and characterization of potential candidates of interest. GeLC-MS/MS is a technically simple method that can resolve large and hydrophobic proteins typically not accessible via 2D gel electrophoresis. In addition, GeLC-MS/MS approach does not necessarily require control samples which could extend its application.

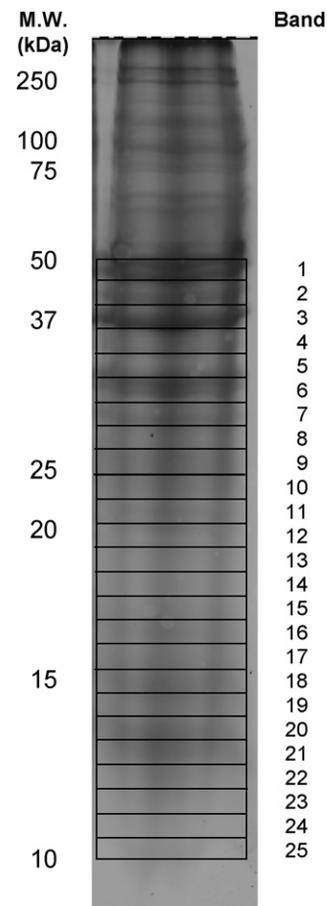


Fig. 7. One-dimensional gel electrophoresis of proteins extracted from virus-infected cells. Boxed areas indicate gel bands excised for mass spectrometric analysis.

5. Conclusions

The data presented demonstrate the applicability of the two proteomics methods, 2D-DIGE and GeLC-MS/MS, for the identification of an infectious agent by expression of viral proteins in cultured cells inoculated with a virus isolate from a sample of an infected patient. Although the techniques have been used for other applications related to virus infection, these published studies focused mainly on the analysis of differential expression of host proteins in response to virus infection, however, not for the detection of viral proteins (for example, Jiang et al., 2005; Dhingra et al., 2007). Up to 8 viral proteins were detected and the identification of protein E3 provides information also for the determination of the specific strain of *vaccinia virus*. The results are consistent with the data derived previously by genome based PCR techniques. In addition to genome-based PCR techniques used commonly for the detection and identification of infectious pathogens, proteomic methods may offer a useful alternative for pathogen discovery. Further refinement and optimization of experimental procedures and conditions of proteomics techniques is in progress to improve sensitivity for detection and efficiency for workloads.

Acknowledgements

The authors would like to thank Inger K. Damon for helpful discussion throughout the preparation of the manuscript. We thank Jakub Baudys for reading and editing of the manuscript.

References

- Aebersold, R., Mann, M., 2003. Mass spectrometry-based proteomics. *Nature* 422, 198–207.
- Anonymous, 2007. Vulvar vaccinia infection after sex with a military smallpox vaccine: case report and use of the pathogen discovery program—Alaska, 2006. *MMWR Morb. Mortal. Wkly. Rep.* 56 (17), 417–419.
- Chong, R.E., Hamler, R.L., Lubman, D.M., Ethier, S.P., Rosenspire, A.J., Miller, F.R., 2001. Differential screening and mass mapping of proteins from premalignant and cancer cell lines using nanoporous reversed-phase HPLC coupled with mass spectrometric analysis. *Anal. Chem.* 73, 1219–1227.
- Chung, C.S., Chen, C.H., Ho, M.Y., Huang, C.Y., Liao, C.L., Chang, W., 2006. Vaccinia virus proteome: identification of proteins in vaccinia virus intracellular mature virion particles. *J. Virol.* 80 (5), 2127–2140.
- Dhingra, V., Li, X., Liu, Y., Fu, Z.F., 2007. Proteomic profiling reveals that rabies virus infection results in differential expression of host proteins involved in ion homeostasis and synaptic physiology in the central nervous system. *J. Neurovirol.* 13 (2), 107–117.
- Gygi, S.P., Rist, B., Cerber, S.A., Turacek, F., Gelb, M.H., Aebersold, R., 1999. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* 17, 994–999.
- Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids Symp. Ser.* 41, 95–98.
- Jiang, X.S., Tang, L.Y., Dai, J., Zhou, H., Li, S.J., Xia, Q.C., Wu, J.R., Zeng, R., 2005. Quantitative analysis of severe acute respiratory syndrome (SARS)-associated coronavirus-infected cells using proteomic approaches: implications for cellular responses to virus infection. *Mol. Cell. Proteomics* 4 (7), 902–913.
- Kang, X., Xu, Y., Wu, X., Liang, Y., Wang, C., Guo, J., Wang, Y., Chen, M., Wu, D., Wang, Y., Bi, S., Qiu, Y., Lu, P., Cheng, J., Xiao, B., Hu, L., Gao, X., Liu, J., Wang, Y., Song, Y., Zhang, L., Suo, F., Chen, T., Huang, Z., Zhao, Y., Lu, H., Pan, C., Tang, H., 2005. Proteomic fingerprints for potential application to early diagnosis of severe acute respiratory syndrome. *Clin. Chem.* 51 (1), 56–64.
- Kuhn, E., Wu, J., Karl, J., Liao, H., Zolg, W., Guild, B., 2004. Quantification of c-reactive protein in the serum of patients with rheumatoid arthritis using multiple reaction monitoring mass spectrometry and ¹³C-labeled peptide standards. *Proteomics* 4, 1175–1186.
- Mann, M., Aebersold, R., 2003. Mass spectrometry-based proteomics. *Nature* 422, 198–207.
- Maxwell, K.L., Frappier, L., 2007. Viral proteomics. *Microbiol. Mol. Biol. Rev.* 398–411.
- O'Farrell, P.H., 1975. High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* 250, 4007–4021.
- Ong, S.E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A., Mann, M., 2002. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* 1, 376–386.
- Resch, W., Hixson, K.K., Moore, R.J., Lipton, M.S., Moss, B., 2007. Protein composition of the vaccinia virus mature virion. *Virology* 358 (1), 233–247.
- Roeperstorff, P., Fohlman, J., 1984. Letter to the editors. *Biomed. Mass Spectrom.* 11, 601.
- Ross, P.L., Huang, Y.L.N., Marchese, J.N., Pappin, D.J.C., 2004. Multiplexed protein quantification in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* 3, 1154–1169.
- Schirle, M., Heurtier, M.-A., Kuster, B., 2003. Profiling core proteomes of human cell lines by one-dimensional page and liquid chromatography-tandem mass spectrometry. *Mol. Cell. Proteomics* 2, 1297–1305.
- Thompson, J.D., Higgins, D.G., Gibson, T.J., Clustal, W., 1994. Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22 (22), 4673–4680.
- Ünlü, M., Morgan, M.E., Minden, J.S., 1997. Differential gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis* 18, 2071–2077.
- Washburn, M.P., Wolters, D., Yates III, J.R., 2001. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* 19, 242–247.
- Wilm, M., Shevchenko, A., Houthaeve, T., Breit, S., Schweigerer, L., Fotsis, T., Mann, M., 1996. Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature* 379, 466–469.
- Wu, J., Lin, Q., Lim, T.K., Liu, T., Hew, C.L., 2007. White spot syndrome virus proteins and differentially expressed host proteins identified in shrimp epithelium by shotgun proteomics and cleavable isotope-coded affinity tag. *J. Virol.* 81 (21), 11681–11689.
- Yao, X., Freas, A., Ramirez, J., Demirev, P.A., Fenselau, C., 2001. Proteolytic ¹⁸O labeling for comparative proteomics: model studies with two serotypes of adenovirus. *Anal. Chem.* 73, 2836–2842.
- Yoder, J.D., Chen, T.S., Gagnier, C.R., Vemulapalli, S., Maier, C.S., Hruby, D.E., 2006. Pox proteomics: mass spectrometry analysis and identification of Vaccinia virion proteins. *Virol. J.* 3 (1), 10.