Defence Research and Development Canada
Recherche et développement pour la défense Canada

DEFENCE R&D DÉFENSE

Mass Spectrometric Analysis of Thiocitrulline Bound Proteins Extracted from Human Keratinocyte Cells

Chemical Proteomics and Mass Spectrometry Experiments to Identify Proteins that Binds to Thiocitrulline

Mulu G. Gebremedhin, David C.W. Mah
Canada West Biosciences Inc., Camrose, AB

Peggy Nelson, Nora W.C. Chan
DRDC Suffield

Defence R&D Canada

Technical Memorandum

DRDC Suffield TM 2008-245

December 2008

Canada

Mass Spectrometric Analysis of Thiocitrulline Bound Proteins Extracted from Human Keratinocyte Cells

Chemical Proteomics and Mass Spectrometry Experiments to Identify Proteins that Binds to Thiocitrulline

Mulu G. Gebremedhin, David C.W. Mah
Canada West Biosciences Inc., Camrose, AB

Peggy Nelson, Nora W.C. Chan
DRDC Suffield

Defence R&D Canada – Suffield

Technical Memorandum
DRDC Suffield TM 2008-245
December 2008

Principal Author

Original signed by Nora W.C. Chan

Nora W.C. Chan

Defence Scientist

Approved by

Original signed by Chris A. Weickert

Chris A. Weickert

Head, Detection and Identification Section

Approved for release by

Original signed by Paul A. D'Agostino

Paul A. D'Agostino

Chair, DRDC Suffield Document Review Panel

Abstract

Previous studies have shown that some arginine analogue compounds, such as L-thiocitrulline, protect human keratinocytes against the toxicity of sulfur mustard (H). Their mechanism of action, however, is not yet known, but in the case of L-thiocitrulline there is reason to suspect that it binds to a cell membrane structure due to its speed of action. In this work we describe the development of proteomics-based approach to identify proteins that bind thiocitrulline using ESI-MS. Membrane proteins were isolated from the first passage of human skin keratinocytes cells which were prepared from neonatal foreskins. Thiocitrulline-bound proteins were subjected to trypsin digestion. Tryptic-digested peptides were eluted and then desalted using SCX zip tip. Peptide separation was done on a C18 capillary column with direct infusion via nanospray ionization to a 3200 QTrap instrument. Independent data acquisition (IDA) files were used with both MASCOT™ 2.2 and ProteinPilot™ 2.0.1 search engines against the Swiss-Prot protein database. A total of 9 proteins were identified using both search engines. Six proteins were suspected of binding to thiocitrulline.

Résumé

Des études ont montré que certains composés analogues à l'arginine, telle la L-thiocitrulline, protègent les kératinocytes humains contre la toxicité de l'ypérite (H). On ne connaît pas encore le mécanisme d'action de ces agents, mais le fait que la L-thiocitrulline agit vite permet de croire qu'elle se lie à une structure de la membrane cellulaire. Dans la présente étude, nous décrivons l'élaboration d'une approche fondée sur la protéomique qui vise à déterminer, au moyen de la méthode d'ES-MS, quelles protéines fixent la thiocitrulline. Des protéines membranaires ont été isolées lors des premiers passages en culture de kératinocytes (cellules de la peau) humains préparés à partir de prépuces de nouveau-nés. Les protéines fixatrices de la thiocitrulline ont été soumises à une digestion par la trypsine. Les peptides ainsi obtenus ont été élués pour ensuite être dessalés au moyen d'embouts de pipettes Zip Tip faits de résine échangeuse de cations forts. La séparation des peptides a été réalisée à l'aide de colonnes capillaires C18 avec infusion directe au moyen de l'ionisation par nanoélectronébulisation sur un appareil 3200 QTrap. Les fichiers d'acquisitions de données indépendantes ont été analysés avec les moteurs de recherche MASCOT^{MC} 2.2 et ProteinPilo^{MC} 2.0.1, en regard des données contenues dans la base sur les protéines Swiss-Prot. Au total, 9 protéines ont été identifiées à l'aide des deux moteurs de recherche. On soupçonne six d'entre elles de fixer la thiocitrulline.

This page intentionally left blank.

Executive Summary

Mass Spectrometric Analysis of Thiocitrulline Bound Proteins Extracted from Human Keratinocyte Cells: Chemical Proteomics and Mass Spectrometry Experiments to Identify Proteins that Binds to Thiocitrulline

Gebremedhin, M.G., Mah, D.C.W., Nelson, P., Chan, N.W.C.; DRDC Suffield TM 2008-245; Defence R&D Canada – Suffield; December 2008.

Background: Although sulfur mustard was first used as a chemical warfare agent over 80 years ago, its biochemical mechanism of action remains unclear and there are no antidotes. Nonetheless, previous work at DRDC Suffield in the 1990s showed that some analogues of the amino acid arginine protect cultured human skin cells against sulfur mustard. The most effective arginine analogue was L-thiocitrulline, a potent and specific inhibitor of the enzyme nitric oxide synthase (NOS), but other NOS inhibitors were shown to be ineffective. Therefore the protection was neither through inhibition of NOS, nor direct interaction between L-thiocitrulline and sulfur mustard. The mechanism of protection remains unclear. However, the rapidity at which protection was afforded suggested that L-thiocitrulline may bind to a cell surface structure rather than an intracellular target.

Since then, several techniques have been developed to identify proteins of interest that are involved in molecular interactions with target molecules. To address the question of the mechanism whereby L-thiocitrulline is effective against sulfur mustard, a new method of identifying cell surface proteins which bind a target molecule, based on chemical proteomics and mass spectrometry, has been developed.

Results: Using this newly-developed method, six proteins derived from human skin cell membranes have been identified which appear to bind L-thiocitrulline.

Significance: The combination of chemical proteomics and mass spectrometry detection could be used to identify the protein cascade involved in thiocitrulline protection against sulfur mustard for better understanding of the pathway involved in toxicity. A better understanding of the pathway involved in sulfur mustard toxicity can be exploited to possibly develop more effective medical countermeasures. The success of this project could lead to a new strategy for the development of effective medical countermeasures to other chemical agents.

Future plans: Future work will include methods development to optimize the affinity chromatography experiments, the extraction reagents and assessment of the utility of two dimensional HPLC separations. Validation of bound protein by methods such as *ex vivo* fluorescent tag binding and confocal microscopy will also be investigated in the future.

Sommaire

Mass Spectrometric Analysis of Thiocitrulline Bound Proteins Extracted from Human Keratinocyte Cells: Chemical Proteomics and Mass Spectrometry Experiments to Identify Proteins that Binds to Thiocitrulline

Gebremedhin, M.G., Mah, D.C.W., Nelson, P., Chan, N.W.C.; DRDC Suffield TM 2008-245; R & D pour la défense Canada – Suffield; Décembre 2008.

Contexte : Même si l'ypérite a été utilisée pour la première fois comme agent de guerre chimique il y a plus de 80 ans, son mécanisme d'action biochimique demeure obscur, et on ne lui connaît aucun antidote. Toutefois, des travaux menés à RDDC Suffield dans les années 1990 ont montré que certains acides aminés analogues à l'arginine protègent les cellules cutanées humaines mises en culture contre l'ypérite. L'analogue de l'arginine le plus efficace était la L-thiocitrulline, un inhibiteur puissant et sélectif de l'enzyme oxyde nitrique synthase (NOS). Par contre, d'autres inhibiteurs du NOS se sont révélés inefficaces. La protection n'était donc assurée ni par l'inhibition du NOS ni par l'interaction directe entre la L-thiocitrulline et l'ypérite. Le mécanisme de protection du composé demeure mal expliqué. Toutefois, la rapidité avec laquelle celui-ci a été enclenché laisse croire que la L-thiocitrulline pourrait se fixer à une structure de la surface des cellules plutôt qu'à une cible intracellulaire.

Depuis, on a conçu plusieurs techniques permettant de trouver des protéines présentant un intérêt en raison de leur potentiel d'interaction avec les molécules ciblées. Afin de comprendre le mécanisme qui explique l'efficacité de la L-thiocitrulline contre l'ypérite, on a élaboré une nouvelle méthode combinant des techniques de protéomique chimique et la spectrométrie de masse, grâce à laquelle il est possible de déterminer les protéines de la surface cellulaire sur lesquelles se fixe une molécule ciblée.

Résultats : En appliquant cette nouvelle méthode, on a repéré six protéines provenant de la membrane de cellules cutanées humaines qui semblent pouvoir fixer la L-thiocitrulline.

Importance : L'emploi combiné de techniques de protéomique chimique et de spectrométrie de masse pourrait servir à déterminer la cascade de protéines entrant en jeu dans la protection conférée par la thiocitrulline contre l'ypérite. Ces techniques permettraient de mieux comprendre le mécanisme de la toxicité de l'ypérite et, de ce fait, de créer des contre-mesures médicales plus efficaces. La réussite du présent projet pourrait mener à la mise sur pied d'une nouvelle stratégie qui serait applicable à l'élaboration de contre-mesures médicales efficaces contre d'autres agents chimiques.

Perspectives : Les travaux à venir porteront notamment sur la conception de méthodes visant à optimiser les expériences de chromatographie d'affinité, sur les réactifs d'extraction et sur l'évaluation de l'utilité de la séparation par CLHP bidimensionnelle. L'identification de protéines fixatrices au moyen de méthodes *ex vivo* telles que le marquage fluorescent des ligands et la microscopie confocale fera également l'objet de recherches futures.

Table of contents

List of Figures

List of Tables

Acknowledgements

The authors would like to thank Dr. Thomas W. Sawyer for his professional advice and kindly proof-reading this document.

Introduction

Sulfur mustard (H) was first used during the First World War in 1917. Since then, there has been nearly a century of H research. Nevertheless, despite the intensive research over the years, the exact cytotoxic mechanism of H remains elusive [1, 2]. Sulfur mustard was historically the most abundantly produced and stockpiled vesicant worldwide, and was used during the Iran - Iraq war (1983-1988). The damaging factor of H is its ability to reduce effectiveness of military personnel with average length of hospitalization of 42 days [1]. There is no antidote, and rapid decontamination is the only effective means of preventing or decreasing tissue damage [1].

L-thiocitrulline (γ-thioureido-L-norvaline, L-TC) is a potent, stereospecific, heme-binding inhibitor of the nitric oxide synthase (NOS) [3]. Sawyer *et al.* reported selected arginine analogues, including L-TC, had a protective effect on the toxicity of H in human keratinocytes [4]. It was demonstrated that the prophylactic protective activity of L-TC was not associated with its interaction with H, nor with its inhibiting potency against NOS, because other NOS inhibitors were ineffective [5 - 7]. Therefore, more information is needed concerning the molecular pathway involved in the protective properties of L-TC. The fast protective action (1 min prior to H exposure) indicates that L-TC may bind to a cell surface structure [6]. Investigation into the identity of the biomolecules that are associated with the thiocitrulline protection against H could lead to a better understanding of the pathway involved in toxicity and development of effective medical countermeasures against H.

Chemical proteomics directly and comprehensively identifies proteins that bind specifically to candidate compounds (in this case, thiocitrulline) by means of affinity chromatographic purification using the immobilized candidate, combined with mass spectrometric identification of interacting proteins [8 - 14]. This is an effective approach for discovering unknown protein functions, identifying the molecular mechanisms of drug action, and obtaining information for optimization of drug lead compounds [10, 11, 14]. Affinity chromatography assays are useful for both confirming the existence of an interaction and as an initial screening assay for identifying previously unknown interactions. Affinity chromatography experiments in proteomics usually rely on biotinylated compounds and subsequent immobilization of these compounds on a streptavidin-functionalized support [8, 9]. Elution of the bound proteins can be the most difficult stage of the affinity chromatography experiments. In general, the ligand-protein interaction is often based on a combination of electrostatic, hydrophobic, and hydrogen bonds. Agents that weaken such interactions might be expected to function as effective eluants. Careful consideration of the relative importance of these three types of interaction will help in the choice of suitable eluants.

In addition to the elution process, detection and identification of limited amounts of bound proteins is one of the challenges in current proteomics studies. Mass spectrometry has become the method of choice for the rapid identification of proteins in proteomic research [12]. In mass spectrometric analysis, biomolecular ions are generated either by electrospray ionization (ESI) or matrix-assisted laser desorption ionization (MALDI). ESI-MS is easily coupled online with various separation techniques such as HPLC. ESI is an atmospheric pressure ionization technique applicable to a wide range of compounds in solution. It generates ions directly from solution (usually in an aqueous/organic solvent system) by creating a fine spray of highly charged droplets in the presence of a strong electric field. As the droplet decreases in size, the electric charge

density on its surface increases. The mutual repulsion between like charges on this surface becomes so great that it exceeds the forces of surface tension, and ions begin to leave the droplet through what is known as the "Taylor cone". The ions are then electrostatically directed into the mass analyzer. Vaporization of these charged droplets results in the production of singly-or multiply-charged gaseous ions. The number of charges retained by an analyte can depend on such factors as the composition and pH of the electrosprayed solvent, as well as the chemical nature of the sample. For peptides and proteins, the degree of protonation is dependent upon the number of basic sites. The ionization process and coupling of ESI to mass spectrometer has been reported in the literature [15-17]. A schematic diagram of the ESI process is shown in Figure 1.

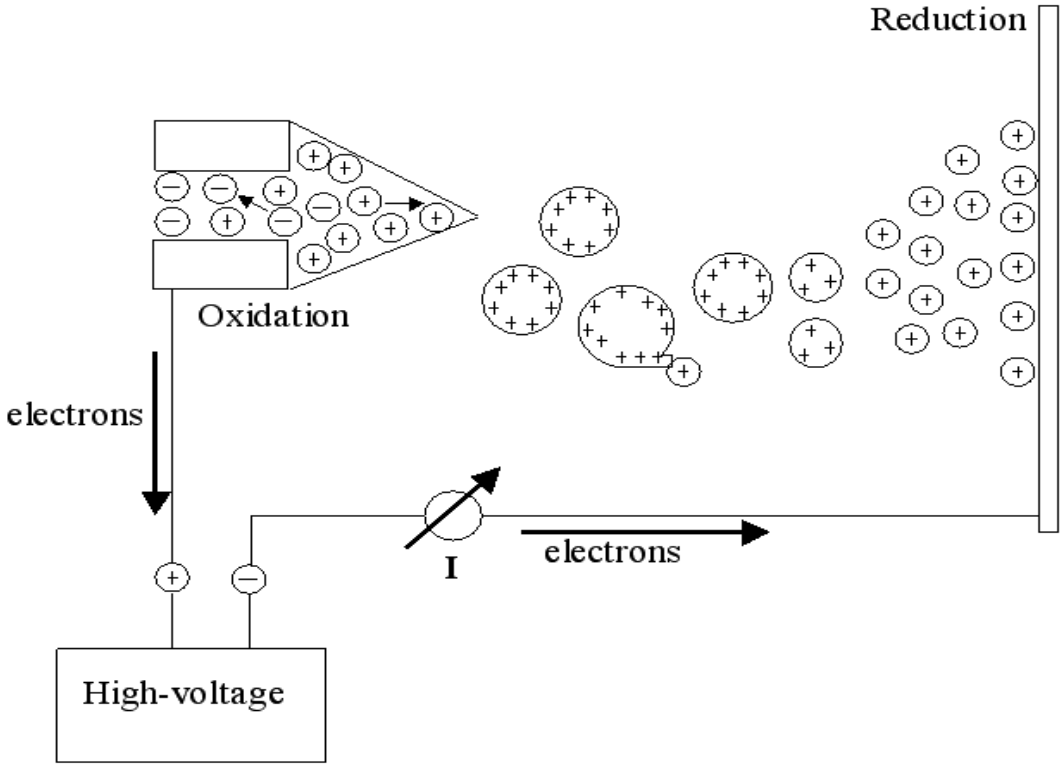

Figure 1: Schematic diagram of the electrospray ionization process.

In most cases, proteins are first digested to peptides by treatment with trypsin. Tryptic-digested samples can be separated on HPLC and detected by MS [18-23], directly infused into MS by static nanoinfusion [24] or nanoinfusion from a syringe pump. In all cases, samples are usually purified by zip tip or other methods before ESI-MS and MS/MS analysis.

Here we describe the first proof-of-concept approach to identify the thiocitrulline-bound proteins based on affinity chromatography experiments and ESI-MS. Membrane proteins were isolated from first passage cultures of human skin keratinocytes cells which were prepared from neonatal foreskins. The proteins were mixed with the beads that contain thiocitrulline. The thiocitrulline-bound proteins were digested with trypsin, and the resulting tryptic-digested peptides were analyzed via tandem mass spectrometry (MS/MS). The MS/MS spectra were acquired with

independent data acquisition methods. A control experiment was also performed to identify any nonspecific binding. ProteinPilot™ (Paragon™ algorithm) and MASCOT™ were used as search engines to identify the bound proteins. ProteinPilot™ 2.0.1 Software [25] streamlines protein identification and quantitation, enabling a search for hundreds of peptide modifications and non-tryptic cleavages simultaneously; easily distinguishes protein isoforms, protein subsets, suppresses false positives; and visualizes peptide-protein associations and relationships. Paragon™ is part of ProteinPilot™ suite of search algorithms and uses sequence tag algorithm to calculate Sequence Temperature Values for identification of peptides from a database. The Paragon™ algorithm enables simultaneous searching for over 150 biological and other modifications, genetic variants and unexpected cleavages without the typical explosion of false positives that plagues traditional algorithms [26]. The MASCOT™ database search engine is a software suite of tools for protein identification that enables mass spectrometry data to be rapidly matched against a FASTA format protein or nucleic acid sequence database. MASCOT™ is a powerful search engine that uses mass spectrometry data to identify proteins from primary sequence databases (MSDB, NCBInr, SwissProt, and dbEst).

Materials and Methods

Reagents

HPLC grade acetonitrile (ACN), HPLC grade water, formic acid (FA, 99+ %), Trypsin Gold, SCX zip tip pipet tips, phosphate buffered saline solution (PBS) and bicinchoninic acid (BCA) reagents were purchased from Thermo Fisher Scientific Inc. (Rockford, IL). Hydrophobic (C18) capillary columns were purchased from Upchurch Scientific, Inc. (Oak Harbor, WA). Keratinocyte Serum Free Media (KSFM) supplemented with bovine pituitary extract and recombinant epidermal growth factor were purchased from Gibco/Invitrogen (Grand Island, NY), dispase was from BD Biosciences (Bedford, MA), and trypsin was from Sigma Aldrich (St.Louis, MO).

Preparation of human keratinocyte cultures

Primary cultures of human skin keratinocytes were prepared from neonatal foreskins [27]. Human foreskin samples were obtained from the Medicine Hat Hospital after infant circumcision. Tissue was obtained on the day of circumcision and placed in dispase overnight; the epidermis was then removed and cells further segregated by incubation in trypsin. After the dispase and trypsin treatments, the cell suspension was centrifuged and keratinocytes resuspended and cultured. Human keratinocytes were grown in keratinocyte-serum-free medium (KSFM) supplemented with 0.005% ($^{w}/_{v}$) bovine pituitary extract and 5 ng/mL recombinant epidermal growth factor in a humidified incubator at 37°C with 5% CO_2. First passage cultures were seeded from log growth primary cultures. The cells were trypsinized and pelleted by centrifugation (~230 × g, 5 min) prior to storage at -70°C. The cell pellet was then washed with PBS buffer and centrifuged at ~230 × g at 4°C for 10 minutes.

Protein extraction

Membrane proteins were extracted using a native membrane protein extraction reagent kit (M-PEK) from Calbiochem (Damstardt, Germany). About 2.61×10^7 cells were treated with M-PEK. The M-PEK kit contains two extraction buffers (I and II). The two buffers were thawed and mixed well by vortexing. During the extraction procedure the extraction buffers were kept on ice and the protease inhibitor at room temperature. An aliquot of protease inhibitor (40 µL) was added to the wall of the tube containing the cell pellet, then 8 mL ice cold buffer I was added immediately to the cell pellet. A pipette was used to mix and resuspend the cell pellet in extraction buffer I. The sample was incubated for 10 min at 4°C under gentle agitation and pelleted at 13,000 × g at 4°C for 20 minutes. The supernatant, enriched with soluble proteins, was removed carefully. Another aliquot of protease inhibitor (5 µL) was added to the wall of the tube, immediately followed by the addition of 4 mL ice cold buffer II. The cell pellet was mixed and resuspended using a pipette and then incubated for 30 minutes at 4°C under gentle agitation. The sample was pelleted at 13,000 × g at 4°C for 20 minutes. The membrane protein extract was aliquoted into 100 µL portions and stored at -20°C. The reagents included in the buffers of M-PEK extraction kit were stated to be compatible with the BCA protein assay, thus the extracts were not subjected to dialysis. The protein concentration of the membrane protein extracts was measured using the enhanced test tube protocol in the BCA protein assay kit.

Protein binding

Thiocitrulline cadaverine biocytin (MW = 702.8, 0.3 μmol), custom synthesized by NAEJA Pharmaceutical Inc. (Edmonton, AB), was immobilized onto 600 μL MPG Streptavidin 30-μm diameter magnetic beads (PureBiotech LLC) for use as a solid phase extraction bed. The immobilization was based on the strong, specific binding between streptavidin and biotin. At a density of 500 pmol streptavidin per μL volume of beads, the ratio of streptavidin to thiocitrulline cadaverine biocytin was 1:1. Three aliquots of membrane proteins (500 μg total) were added to the thiocitrulline MPG beads. After each addition of membrane protein, the bead-protein mixture was vortexed and incubated at room temperature with agitation for 30 minutes, then the supernatant was removed from the beads using a magnetic separator. Finally, the beads were washed three times with lysis buffer and then with ammonium bicarbonate buffer (50 mM). A control experiment of membrane-bound proteins mixed with magnetic beads (without thiocitrulline) was also performed to identify any non-specific binding.

Trypsin digestion and elution of bound proteins

After multiple washings, 20 μL of 50 mM ammonium bicarbonate was added and the beads containing the bound proteins were heated to 95°C for 5 minutes to denature the proteins. Trypsin Gold (0.4 μg) was added and the mixture was incubated at 37°C overnight. After collecting the supernatant from the beads, 20 μL of 0.1% FA in 30% ACN/H_2O was added to the beads to further extract the peptides.

Peptide separation and MS/MS analysis

All the solution (supernatant and FA/ACN/H_2O extract) was pooled together and vacuum-dried in a speed-vac concentrator. The dried sample was reconstituted with 0.1% trifluoroacetic acid (TFA) followed by a desalting step using a SCX zip tip. Peptide separations were performed on a C18 packed fused-silica capillary column (25 mm × 0.1 mm i.d.). The tryptic-digested mixture was loaded into the C18 column by direct infusion using a syringe pump (Harvard PicoPlus) at a flow rate of 0.3 μL/min. A solution of 0.1% FA in 2% ACN/H_2O was infused into the column for 5 minutes and then followed by a solvent step gradient: 0.1% FA in 5% ACN/ H_2O for 5 minutes, 0.1% FA in 10% ACN/ H_2O for 5 minutes , 0.1% FA in 20% ACN/ H_2O for 5 minutes, 0.1% FA in 30% ACN/ H_2O for 10 minutes, 0.1% FA in 50% ACN/ H_2O for 10 minutes, and finally 0.1% FA in 95% ACN/ H_2O for another 10 minutes. IDA (information-dependent acquisition) data were acquired using an Applied Biosystems 3200 QTrap interfaced with a nanospray source. The IDA method created included an IDA criteria (specify the charge state, mass range and rolling collision energy), enhanced MS scan, enhanced resolution scan (to confirm charge state and/or isotope pattern selection), enhanced product ion scan or MS/MS scan.

Database search

IDA files were used for a database search using ProteinPilot™ Paragon™ algorithm. All matches above a 95% confidence interval were considered. To eliminate false positive proteins when a species specific search was conducted, a FASTA file containing common contaminant proteins was created based on the procedures included in the software. The presence of "contaminant" in the species field in this file instructs the Paragon™ algorithm to search contaminant proteins, irrespective of the species setting in the method. An example of the Paragon™ method creation window is shown in Figure 2. Knowledge of the experimental details of the sample is required for input into the Paragon™ method. In this method there are two search efforts, Rapid and Thorough ID. Rapid ID generates fast results similar to standard search engines such as MASCOT™ search. However, the Thorough ID searches consider any modifications selected in the ID Focus pull down menu, including those of much lower probability. All cleavage variants were searched, including those that did not conform to the selected digest agent. The same IDA files were also used in a MASCOT™ ver2.2 database search against the Swiss-Prot database.

The precision tolerance for ESI-MS/MS data was 1.2 Da for peptide masses and 0.6 Da for fragment ion masses. The peptide hits were considered as positive identifications if the MASCOT™ search scores were higher than the MASCOT™ identity scores on a 95% confidence level.

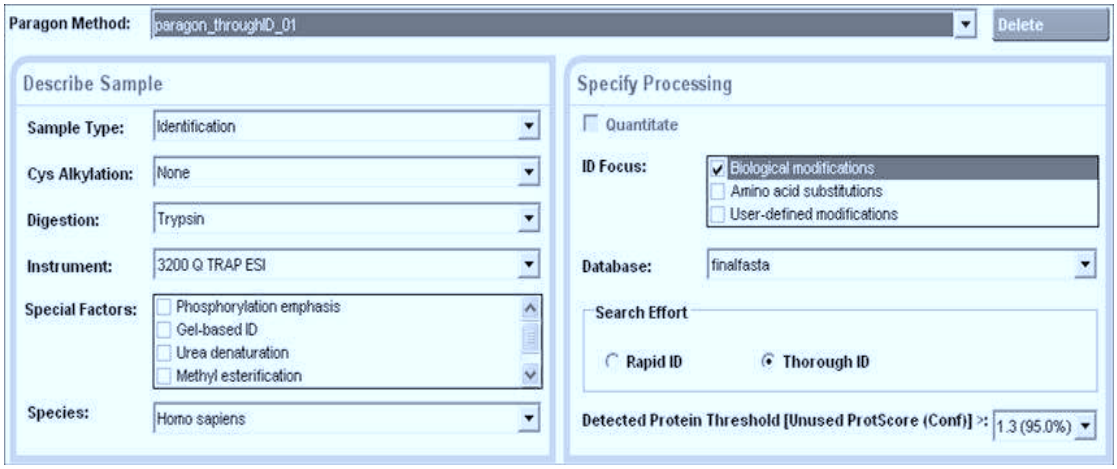

Figure 2: An example of a Paragon™ method creation window.

The chemical proteomics experiment with liquid chromatography separation and mass spectrometry detection is summarized in a flow chart in Figure 3.

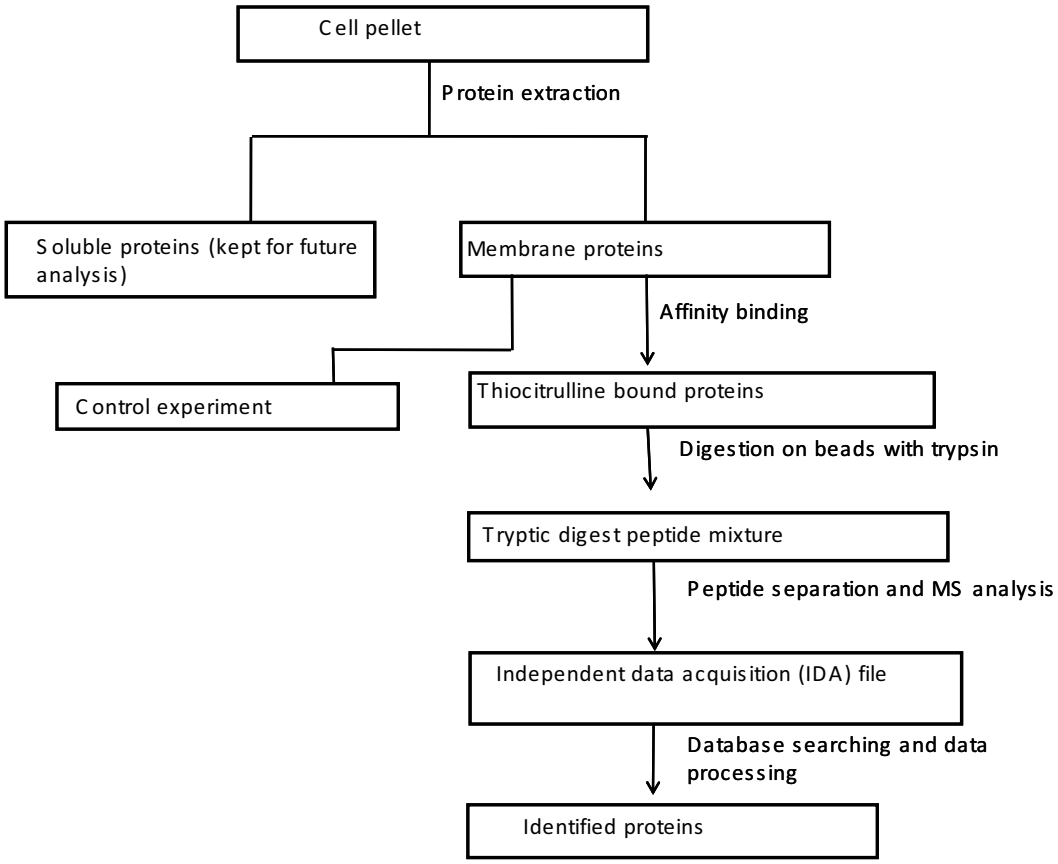

Figure 3: A flow chart of the chemical proteomics and LC-MS process.

Results and Discussion

Protein concentration

The protein concentration standard curve was obtained using serially diluted bovine serum albumin (BSA) in the bicinchoninic acid (BCA) protein assay kit, Figure 4. The total protein concentration in the membrane protein extract from the M-PEK kit was calculated from the standard curve. The diluted membrane protein sample was measured in triplicates and the average absorbance was 0.404 AU at 562 nm. The total protein concentration was estimated at 1.04 µg/µL. The total amount of membrane proteins was 3.17 mg in 2.61×10^7 cells.

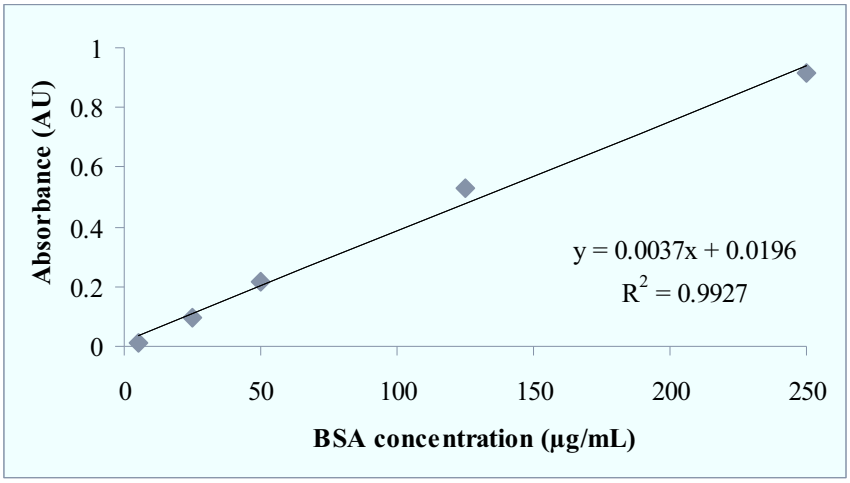

Figure 4: A typical BCA standard curve. The standard curve was obtained using serially diluted bovine serum albumin.

Peptide separation and MS/MS spectra

The thiocitrulline-bound proteins were identified from two independent automated information-dependent acquisition (IDA) experiments. One of the total ion chromatograms (TIC) from the two IDA data is shown on Figure 5. The database search method compares experimental MS/MS spectra with theoretical MS/MS spectra of each peptide derived from protein sequence databases and reports the best match. The primary peptide sequence, amount of internal energy and charge state are among the factors that affect the type of fragment ions observed in an MS/MS spectrum. The ions observed in an MS/MS spectrum are either a, b or c if the charge retains on the N-terminal fragment. If the charge retains on the C-terminal fragment, the ions are classified as either x, y or z. A subscript in the type of ion indicates the number of residues in the fragment. Among the ion types mentioned, b- and y-ions are often observed in low collisionally induced dissociation (CID) energy. Doubly charged ions usually have charges at the opposite ends of the peptide. When these ions are fragmented, both b-ion and the corresponding y-ion are formed. However, when singly charged ions are fragmented, either a b-ion or a y-ion is formed. The other part of the peptide is lost as a neutral fragment. The MS/MS spectra of the matched peptides were examined manually.

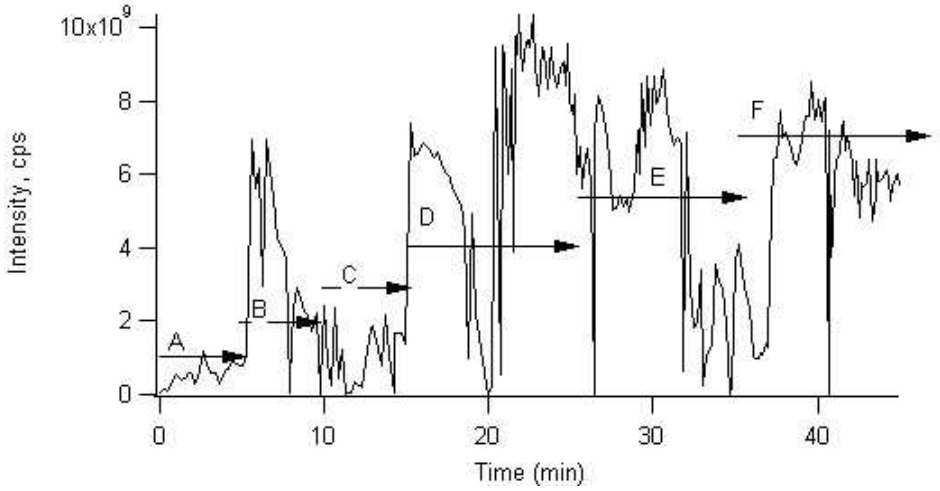

Figure 5: Total ion chromatogram of tryptic-digested peptides derived from proteins bound to thiocitrulline. The solvent step gradient is represented by arrows: 5% B for 5 minutes (A), 10% B for 5 minutes (B), 20% B for 5 minutes (C), 30% B for 10 minutes (D), 50% B for 10 minutes (E) and 95% B for 10 minutes (F). The composition of B was 0.1% FA in ACN.

Some MS/MS spectra of selected peptides including the theoretical sequence ions of the peptide are shown in Figures 6-8. These selected peptides were eluted in region B (m/z 839.6, 5.72 min), region B (m/z 683.3, 5.82 min), and region C (m/z 713.5, 10.12 min) of the above TIC. All the selected peptides shown are doubly charged ions.

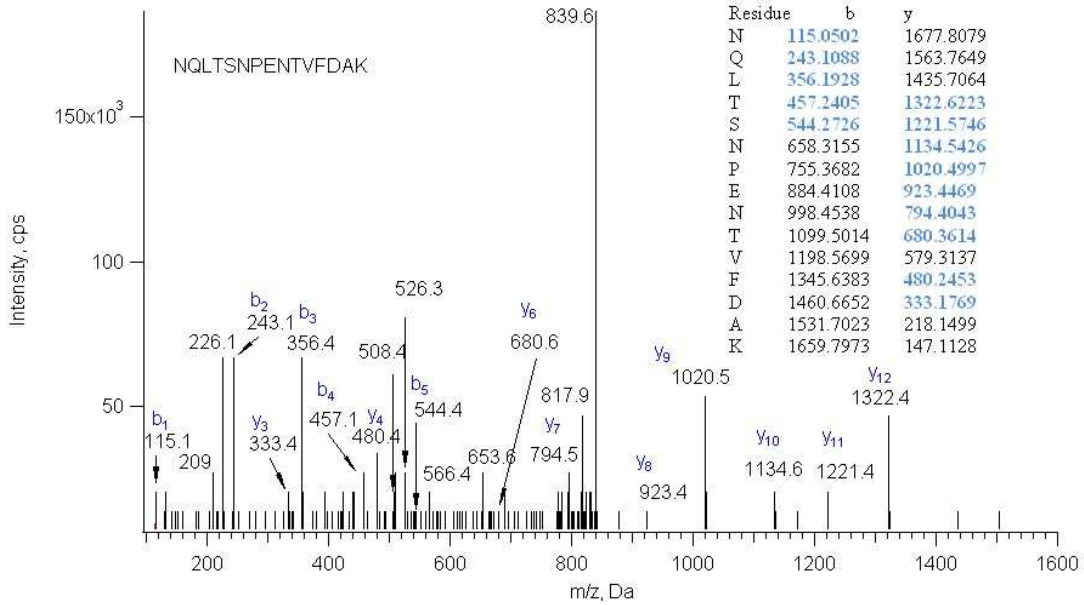

Figure 6: Experimental MS/MS spectrum of peptide NQLTSNPENTVFDAK and theoretical sequence ions of the same peptide (insert, matching ions in blue text) are shown. The peptide represents the protein with accession # of P11021. The parent m/z value of the peptide ion is 839.6

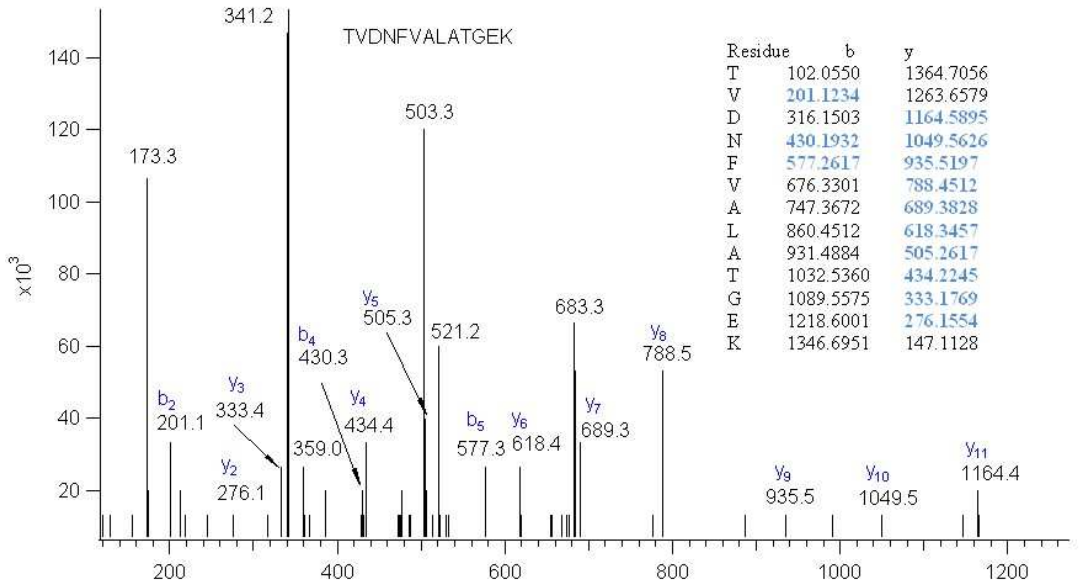

Figure 7: Experimental MS/MS spectrum of peptide TVDNFVALATGEK and theoretical sequence ions of the same peptide (insert, matching ions in blue text) are shown. The peptide represents the protein with accession # of P23284. The parent m/z value of the peptide ion is 683.3.

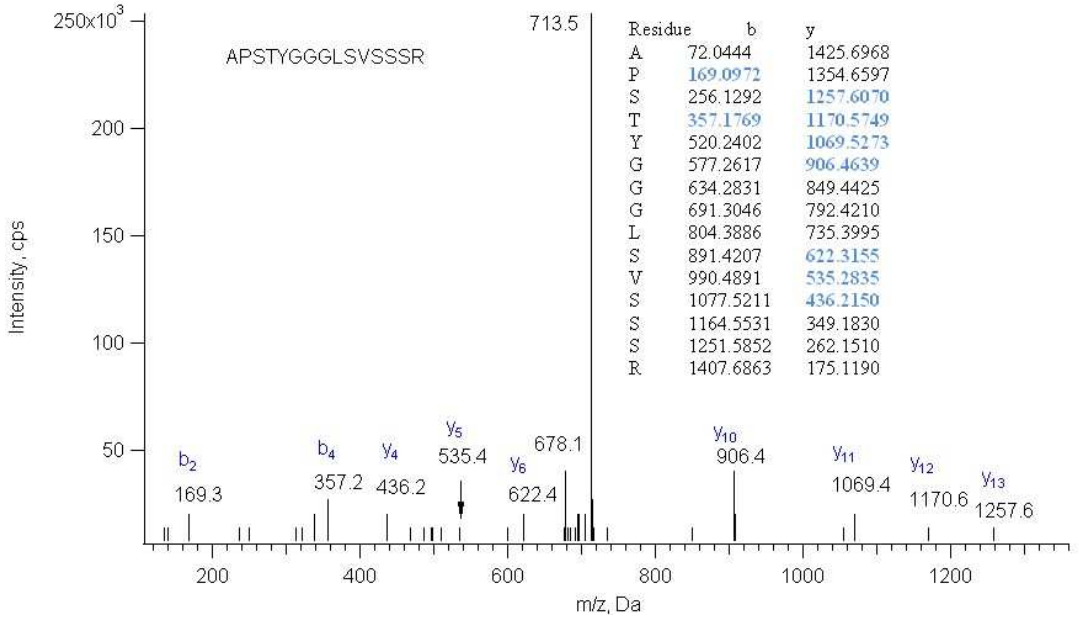

Figure 8: Experimental MS/MS spectrum of peptide APSTYGGGLSVSSSR and theoretical sequence ions of the same peptide (insert, matching ions in blue text) are shown. The peptide represents protein with accession # of P02533. The parent m/z value of the peptide ion is 713.5

Database search and processing

Each IDA method was created manually. In an IDA experiment, an MS survey scan was used to generate a peak list of all ions present. The ion peak list was subjected to a set of user defined criteria to filter out unwanted precursor ions. The remaining ions were then submitted for tandem mass spectrometry (MS/MS). This cycle was repeated throughout the duration of the acquisition to generate large amounts of informative data. IDA files from the control samples were searched using the ProteinPilot™ Paragon™ algorithm in the Thorough ID mode against the Uniprot version of the Swiss-Prot database (with "contaminant proteins" added). The search of the control sample yielded 2 possible proteins identified with scores above 95% confidence threshold, as shown in Table 1. The table includes percentage of sequence coverage which refers to the percentage of matching amino acids from identified peptides having confidence greater than 0 divided by the total number of amino acids in the sequence and "Unused ProtScore" which is a measure of the protein confidence for a detected protein, calculated from the peptide confidence for peptides from spectra that have not already been completely "used" by higher scoring winning proteins. For improved protein identification confidence, the same IDA file was searched using MASCOT™ ver2.2. Vimentin was also identified from the MASCOT™ search (data not shown).

Similarly, IDA files from sample containing proteins bound to thiocitrulline were generated. Eight proteins associated with thiocitrulline were found using the ProteinPilot Paragon™ search with scores above 95% confidence, as shown in Table 2. Two of the identified proteins were identified as contaminants, serum albumin and keratin 1.

Table 1: List of proteins identified by ProteinPilot (Paragon algorithm) from a control experiment.

Protein identified	**Accession #**	**Sequence coverage (%)**	**Unused ProtScore**
Vimentin	P08670	6.22	2.96
Activating signal cointegrator 1 complex subunit 3	Q8N3C0	0.408719	1.40

Table 2: List of proteins identified by ProteinPilot (Paragon algorithm) from a thiocitrulline treated experiment.

Protein identified	**Accession #**	**Sequence coverage (%)**	**Unused ProtScore**
Serum albumin precursor (contaminant)	cont|000097	42.7	4.07
Keratin, type II cytoskeletal 1 Keratin 1 (contaminant)	cont|000135	45.3	2.98
WD repeat and FYVE domain-containing protein 3	Q8IZQ1	15.3	2.77
Vimentin	P08670	43.1	2.52
78 kDa glucose-regulated protein precursor (GRP 78)	P11021	35.0	2.32
Peptidyl-prolyl cis-trans isomerase B precursor	P23284	26.0	2.23
Keratin, type I cytoskeletal 14	P02533	47.0	2.0
Histone H4	P62805	75.5	2.0

Tables 3 and 4 summarize the results of the Paragon^{TM} search including the number of detected proteins, the number of distinct peptides associated with each of the identified proteins, and the number of spectra used to positively identify these peptides. In both control and thiocitrulline-bound protein experiments, the mass spectral data were searched against 16422 proteins, the numbers of protein detected, with 95% confidence, were 2 and 8 respectively.

Table 3: Overview of the total results from the control experiment in Paragon^{TM} search, the summary table displays the number of proteins detected, the number of distinct peptides identified for these proteins, and the number of spectra used to identify these peptides. The cutoff-threshold was applied at the 95% confidence level (in red).

ID Statistics (Protein-Thresholded): 753 total spectra, 580 non-empty spectra					
Unused (% Confidence) Cutoff	Proteins Detected	Proteins Before Grouping	Distinct Peptides	Spectra Identified	% Total Spectra
>2.0(99)	1	1	2	4	0.5
>1.3 (95)	2	3	3	5	0.7
>0.47 (66)	4	6	7	11	1.5

Table 4: Overview of the total results from the thiocitrulline-bound protein identification experiment, the summary table displays the number of proteins detected, the number of distinct peptides identified for these proteins, and the number of spectra used to identify these peptides. The cutoff-threshold was applied at the 95% confidence level (in red).

ID Statistics (Protein-Thresholded): 5165 total spectra, 4266 non-empty spectra					
Unused (% Confidence) Cutoff	Proteins Detected	Proteins Before Grouping	Distinct Peptides	Spectra Identified	% Total Spectra
>2.0(99)	6	11	110	122	2.4
>1.3 (95)	8	13	124	137	2.7
>0.47 (66)	9	15	136	149	2.9

Six of the 8 proteins found using the Paragon™ algorithm search were also confirmed by MASCOT™. However, one other protein (thrombospondin-2 precursor) was found only from the MASCOT™ search, as summarized in the Venn diagram in Figure 9. Overall, six thiocitrulline-bound proteins were identified using both search engines, where vimentin was a result of non-specific binding and serum albumin and keratin 1 were contaminants.

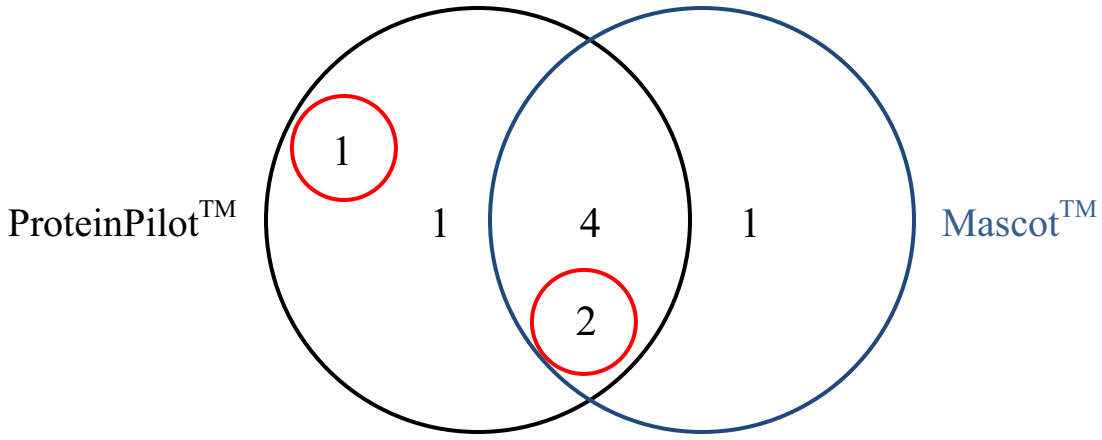

Figure 9: A Venn diagram showing the number of proteins obtained from ProteinPilot™ and MASCOT™ search engines for thiocitrulline treated experiment. The values in red circles represent proteins identified in either the control experiment or as contaminants.

Conclusion

A chemical proteomics and LC-MS approach based on compound-immobilized affinity chromatography protocol was developed to identify thiocitrulline-bound proteins extracted from human keratinocyte cells. IDA files were searched against Swiss-Prot database using ProteinPilot™ and MASCOT™ search engines. Overall, 9 proteins were detected. Six of the 9 proteins were found by both search engines, two proteins were only found from the Paragon™ search, and one unique protein was found only from the MASCOT™ search. Two of the 9, serum albumin and keratin, were contaminant proteins. A control experiment was also performed and identified 2 proteins, vimentin and activating signal cointegrator 1, as non-specific binders. Therefore, six proteins were identified as thiocitrulline-bound proteins. Both ProteinPilot™ and MASCOT™ search engines are needed to ensure high confidence in protein identification.

This body of work is a first proof of concept to develop a chemical proteomics and LC-MS method to identify thiocitrulline-bound proteins. Future work will include further method development to optimize the affinity capture experiments, the extraction reagents and assessment of the utility of two dimensional HPLC separations. Validation of bound protein by methods such as *ex vivo* fluorescent tag binding and confocal microscopy will also be investigated in the future.

References

[1] Hurst, C.G., Petrali, J.P., Barillo, D.J., Graham, J.S., Smith, W.J., Urbanetti, J.S., and Sidell, F.R. (2008). Vesicants. In Lenhart, M.K. (ed.) *Medical aspects of chemical warfare*, pp.259-310. Office of The Surgeon General at TMM Publications, Borden Institute, Walter Reed Army Medical Center, Washington, DC. http://www.bordeninstitute.army.mil/published_volumes/chemwarfare/CH8_Pgs259-310.pdf (18 August 2010)

[2] Kehe, K., Balszuweit, F., Emmler, J., Kreppel, H., Jochum, M., and Thiermann, H. (2008). Sulfur mustard research – strategies for the development of improved medical therapy. *ePlasty*, 8, 312-332. http://www.eplasty.com/images/PDF/eplasty08e32.pdf (18 August 2010)

[3] Frey, C., Narayanan, K., McMillan, K., Spack, L., Gross, S.S., Masters, B.S., and Griffith, O.W. (1994). L-Thiocitrulline. A stereospecific, heme-binding inhibitor of nitric-oxide synthese. *Journal of Biological Chemistry*, 269(42), 26083-26091.

[4] Sawyer, T.W. (1999). Synergistic protective effects of selected arginine analogues against sulphur mustard toxicity in neuron culture. *Toxicology and Applied Pharmacology*, 155(2), 169-176.

[5] Sawyer, T.W., Hancock, J.R., and D'Agostino, P.A. (1998). L-Thiocitrulline: a potent protective agent against the toxicity of sulfur mustard in vitro, *Toxicology and Applied Pharmacology*, 151, 340–346.

[6] Sawyer, T.W., and Risk, D. (2000). Effects of selected arginine analogues on sulphur mustard toxicity in human and hairless guinea pig skin keratinocytes. *Toxicology and Applied Pharmacology*, 163(1), 75-85.

[7] Rappeneau, S., Baeza-Squiban, A., Marano, F., and Calvet, J. (2000). Efficient protection of human bronchial epithelial cells against sulphur and nitrogen mustard cytotoxicity using drug combinations. *Toxicological Sciences*, 58(1), 153-160.

[8] Schulze, W. X., and Mann, M. (2004). A novel proteomic screen for peptide-protein interactions. *Journal of Biological Chemistry*, 279, 10756 –10764

[9] Wilhelmsen, K., Copp. J., Glenn, G., Hoffman, R.C., Tucker, P., and Van der Geer, P. (2004). Purification and identification of protein-tyrosine kinase-binding proteins using synthetic phosphopeptides as affinity reagents. *Molecular & Cellular Proteomics*, 3, 887-895.

[10] Katayama, H., and Oda, Y. (2007) Chemical proteomics for drug discovery based on compound-immobilized affinity chromatography. *Journal Chromatography B, Analytical Technolgies in the Biomedical and Life Sciences*. 855(1), 21-27.

[11] Kang, H.J., Yoon, T-S., Jeong, D.G., Kin, Y., Chung, J.W., Ha, J.S., Park, S.S., Ryu, S.E., Kim, S., Bae, K.H., and Chung, S.J. (2008). Identification of proteins binding to decursinol by chemical proteomics. *Journal of Microbiology and Biotechnology*, 18(8), 1427-1430.

[12] Cravatt, B.F., Simon, G.M., and Yates, J.R. III. (2007). The biological impact of mass-spectrometry-based proteomics. *Nature*, 450(13), 991-1000.

[13] Salisbury, C.M., and Cravatt, B.F. (2007). Activity-based probes for proteomic profiling of histone deacetylase complexes. *Proceedings of the National Academy of Sciences*, 104(4), 1171-1176.

[14] Verhelst, S.H.L., and Bogyo M. (2005). Chemical proteomics applied in target identification and drug discovery. *BioTechniques*, 38(2), 1-3.

[15] Chelius, D., Zhang, T., Wang, G., and Shen, R. (2003). Global Protein Identification and Quantification Technology Using Two-Dimensional Liquid Chromatography Nanospray Mass Spectrometry. *Analytical Chemistry*, 75, 6658-6665.

[16] Shevchenko, A., Jensen, O.N., Podtelejnikov, A.V., Sagliocco, F., Wilm, M., Vorm, O., Mortensen, P., Shevchenko, A, Boucherie, H., and Mann, M.(1996). *Proceedings of the National Academy of Sciences of the United States of America*, 93, 14440-14445.

[17] Link, A.J., Eng, J., Schieltz, D.M., Carmack, E., Mize, G.J., Morris, D.R., Garvik, B.M., and Yates, J.R.III (1999). Direct analysis of protein complexes using mass spectrometry. *Nature Biotechnology*, 17, 676-682.

[18] Choi, J., Levey, A., Weintraub, S.T., Rees, H.D., Geraing, M., Chin, L.-S., and Li, L. (2004). Oxidative modifications and down-regulation of ubiquitin carboxyl-terminal hydrolase L1 associated with idiopathic Parkinson's and Alzheimer's diseases. *Journal of Biological Chemistry*, 279(13), 13256-13264.

[19] Cooper, B., Eckert, D., Andon, N.L., Yates, J.R.III, and Haynes, P.A. (2003). Investigative proteomics: identification of an unknown plant virus from infected plants using mass spectrometry. *Journal of American Society for Mass Spectrometry*, 14(7), 736-741.

[20] Drummelsmith, J., Girard, I., Trudel, N., and Ouellette, M. (2004). Differential protein expression analysis of *Leishmania major* reveals novel roles for methionine adenosyltransferase and S-adenosylmethionine in methotrexate resistance. *Journal of Biological Chemistry*, 279(32), 33273-33280.

[21] Schiltz, S., Gallardo, K., Huart, M., Negroni, L., Sommerer, N., and Burstin, J. (2004). Proteome reference maps of vegetative tissues in pea. An investigation of nitrogen mobilization from leaves during seed filling. *Plant Physiology*, 135(4), 2241-2260.

[22] Watson, B. S., Lei, Z., Dixon, R. A., and Sumner, L. W. (2004). Proteomics of *Medicago sativa* cell walls. *Phytochemistry*, 65, 1709–1720.

[23] Han, D. K., Eng, J., Zhou, H., and Aebersold, R. (2001). Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. *Nature Biotechnology*, 19, 946–951.

[24] Chen. S. (2006). Rapid protein identification using direct infusion nanoelectrospray ionization mass spectrometry. *Proteomics,* 6, 16-25.

[25] Applied Biosystems Technical Note # 114TN15-01 (2006) ProteinPilot™ software for protein identification and expression analysis. http://www3.appliedbiosystems.com/cms/groups/psm_marketing/documents/generaldocuments/cms_042621.pdf (18 August 2010)

[26] Shilov, I.V., Seymour, S.L., Patel, A.A., Loboda, A., Tang, W.H., Keating, S.P., Hunter, C.L., Nuwaysir, L.M., and Schaeffer, D.A. (2007) The Paragon algorithm: a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra. *Molecular & Cellular Proteomics*, 6, 1638-1655.

[27] Sawyer, T.W., and Hamilton, M.G. (2000). Effect of intracellular calcium modulation on sulphur mustard cytotoxicity in cultured human neonatal keratinocytes. *Toxicology In Vitro*, 14(2), 149-157.

List of symbols/abbreviations/acronyms/initialisms

ACN	Acetonitrile
AU	Absorbance unit
BCA	Bicinchoninic acid
CID	Collisionally induced dissociation
DRDC	Defence Research & Development - Canada
ESI	Electrospray ionization
FA	Formic acid
H	Sulfur mustard
HPLC	High performance liquid chromatography
i.d.	Inner diameter
IDA	Independent data acquisition
kDa	Kilodalton
m/z	Mass-to-charge
MALDI	Matrix-assisted laser desorption ionization
MS/MS	Tandem mass spectrometry
MW	Molecular weight
rpm	Revolution per minute
SCX	Strong cation exchange
TFA	Trifluoroacetic acid
$^{w}/_{v}$	Weight per volume

Glossary

MSDB

a comprehensive, non-identical protein sequence database maintained by the Proteomics Department at the Hammersmith Campus of Imperial College London. MSDB is designed specifically for mass spectrometry applications.

NCBInr

a comprehensive, non-identical protein database maintained by NCBI for use with their search tools BLAST and Entrez. The entries have been compiled from GenBank CDS translations, PIR, SWISS-PROT, PRF, and PDB.

SwissProt

a high quality, curated protein database. Sequences are non-redundant, rather than non-identical, so you may get fewer matches for an MS/MS search than you would from a comprehensive database, such as MSDB or NCBInr. SwissProt is ideal for peptide mass fingerprint searches.

dbEST

the division of GenBank that contains "single-pass" cDNA sequences, or Expressed Sequence Tags, from a number of organisms. During a Mascot search, the nucleic acid sequences are translated in all six reading frames. dbEST is a very large database, and is divided into three sections: EST_human, EST_mouse, and EST_others. Even so, searches of these databases take far longer than a search of one of the non-redundant protein databases. You should only search an EST database if a search of a protein database has failed to find a match.

Sequence Temperature Values

Sequence Temperature Values are computed using a sequence tag algorithm, allowing the degree of implication by an MS/MS spectrum of each region of a database to be determined on a continuum. Counter to conventional approaches, features such as modifications, substitutions, and cleavage events are modeled with probabilities rather than by discrete user-controlled settings to consider or not consider a feature. The use of feature probabilities in conjunction with Sequence Temperature Values allows for a very large increase in the effective search space with only a very small increase in the actual number of hypotheses that must be scored.

This page intentionally left blank.

DOCUMENT CONTROL DATA

(Security classification of title, body of abstract and indexing annotation must be entered when the overall document is classified)

1. ORIGINATOR (The name and address of the organization preparing the document. Organizations for whom the document was prepared, e.g. Centre sponsoring a contractor's report, or tasking agency, are entered in section 8.)

Defence R&D Canada – Suffield
P.O. Box 4000, Station Main
Medicine Hat, Alberta T1A 8K6

2. SECURITY CLASSIFICATION (Overall security classification of the document including special warning terms if applicable.)

UNCLASSIFIED

3. TITLE (The complete document title as indicated on the title page. Its classification should be indicated by the appropriate abbreviation (S, C or U) in parentheses after the title.)

Mass Spectrometric Analysis of Thiocitrulline Bound Proteins Extracted from Human Keratinocyte Cells: Chemical Proteomics and Mass Spectrometry Experiments to Identify Proteins that Binds to Thiocitrulline

4. AUTHORS (last name, followed by initials – ranks, titles, etc. not to be used)

Gebremedhin, M.G., Mah, D.C.W., Nelson, P., Chan, N.W.C.

5. DATE OF PUBLICATION (Month and year of publication of document.)	6a. NO. OF PAGES (Total containing information, including Annexes, Appendices, etc.)	6b. NO. OF REFS (Total cited in document.)
December 2008	32	27

7. DESCRIPTIVE NOTES (The category of the document, e.g. technical report, technical note or memorandum. If appropriate, enter the type of report, e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.)

Technical Memorandum

8. SPONSORING ACTIVITY (The name of the department project office or laboratory sponsoring the research and development – include address.)

Defence R&D Canada – Suffield
P.O. Box 4000, Station Main
Medicine Hat, Alberta T1A 8K6

9a. PROJECT OR GRANT NO. (If appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant.)	9b. CONTRACT NO. (If appropriate, the applicable number under which the document was written.)

10a. ORIGINATOR'S DOCUMENT NUMBER (The official document number by which the document is identified by the originating activity. This number must be unique to this document.)	10b. OTHER DOCUMENT NO(s). (Any other numbers which may be assigned this document either by the originator or by the sponsor.)
DRDC Suffield TM 2008-245	

11. DOCUMENT AVAILABILITY (Any limitations on further dissemination of the document, other than those imposed by security classification.)

Unlimited

12. DOCUMENT ANNOUNCEMENT (Any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availability (11). However, where further distribution (beyond the audience specified in (11) is possible, a wider announcement audience may be selected.))

Unlimited

13. ABSTRACT (A brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), (R), or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual.)

Previous studies have shown that some arginine analogue compounds, such as L-thiocitrulline, protect human keratinocytes against the toxicity of sulfur mustard (H). Their mechanism of action, however, is not yet known, but in the case of L-thiocitrulline there is reason to suspect that it binds to a cell membrane structure due to its speed of action. In this work we describe the development of proteomics-based approach to identify proteins that bind thiocitrulline using ESI-MS. Membrane proteins were isolated from the first passage of human skin keratinocytes cells which were prepared from neonatal foreskins. Thiocitrulline-bound proteins were subjected to trypsin digestion. Tryptic-digested peptides were eluted and then desalted using SCX zip tip. Peptide separation was done on a C18 capillary column with direct infusion via nanospray ionization to a 3200 QTrap instrument. Independent data acquisition (IDA) files were used with both MASCOT™ 2.2 and ProteinPilot™ 2.0.1 search engines against the Swiss-Prot protein database. A total of 9 proteins were identified using both search engines. Six proteins were suspected of binding to thiocitrulline.

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (Technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus, e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus identified. If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

chemical proteomics, mass spectrometry, HPLC, sulfur mustard, thiocitrulline

Defence R&D Canada

Canada's Leader in Defence
and National Security
Science and Technology

R & D pour la défense Canada

Chef de file au Canada en matière
de science et de technologie pour
la défense et la sécurité nationale

www.drdc-rddc.gc.ca