

The Waters Protein Expression System for Qualitative and Quantitative Proteomics

Therese McKenna, Iain Campuzano, Mark Ritchie, Paul B. Young, Scott J. Geromanos, Jeffrey C. Silva, Frank W. Riley, Kiernan Neeson, Richard Denny, Keith Richardson, James I. Langridge

Waters Corporation

This is an Application Brief and does not contain a detailed Experimental section.

For research use only. Not for use in diagnostic procedures.

Abstract

This application brief describes the implementation of a novel multiplexed approach to proteomics that is capable of both qualitative protein identification, and relative quantification across sample sets. The methodology is based upon an exact mass LC-MS experiment on a Q-ToF mass spectrometer. During the course of an LC-MS run the Q-ToF is programmed in the MS mode to cycle between low and elevated collision energy acquisitions. The resolution and mass measurement accuracy (<5 ppm RMS) obtained from the Q-ToF is key in providing additional specificity for the analysis of complex tryptic digests, allowing confident protein identification.

Benefits

This is a global approach that is ideal for the study of protein expression and the discovery and characterization of biomarkers

Introduction

The Waters Protein Expression System sets the standard for multiplexed qualitative and quantitative proteomics. It combines the highly reproducible nanoACQUITY UltraPerformance LC (UPLC) System, with the specificity of exact mass measurement from the Waters Micromass Q-ToF Premier Mass Spectrometer. Integrated with proprietary Protein Expression System Informatics Tools, this complete solution provides enhanced qualitative protein identification simultaneously with relative quantification across sample sets, without the need for stable isotope labelling. This allows both the identification and characterization of potential biomarkers.



The Waters Protein Expression System is setting new standards in qualitative and quantitative proteomics.

The completion of the Human Genome Project has allowed the rapid identification of large numbers of human proteins using mass spectrometry in combination with bioinformatics. The separation of complex protein mixtures in large proteomic studies has, most commonly, been achieved by 2D-PAGE.

However, recent advances in both HPLC and mass spectrometry instrumentation have allowed the analysis of protein complexes which have not been separated on a two-dimensional gel. These experiments involve separation of the complex digest mixture by micro-capillary liquid chromatography

connected to an instrument capable of data directed analysis (DDA). In this case the mass spectrometer will identify the peptide ions as they elute from the HPLC column and select each precursor ion individually for MS/MS. This makes DDA-enabled MS/MS a serial process, independent of which ionization technique or mass analyzer is used. Protein identification is then achieved via databank searching of the ESI-MS/MS data, providing qualitative information on the proteins that are present. Hundreds of MS/MS spectra can be acquired in a fully automated fashion, resulting in the identification of significant numbers of proteins, including low copy number proteins, from a single LC-MS/MS experiment.

A common goal of these experiments is the qualitative identification of proteins from simple or complex mixtures of biological origin. However, a significant problem in either the gel-based or non-gel based approach described above is the ability to compare relative expression levels of identical proteins between samples. This requires excellent chromatographic reproducibility, selectivity and specificity.

This application brief describes the implementation of a novel multiplexed approach to proteomics that is capable of both qualitative protein identification, and relative quantification across sample sets. The methodology is based upon an exact mass LC-MS experiment on a Q-ToF mass spectrometer. During the course of an LC-MS run the Q-ToF is programmed in the MS mode to cycle between low and elevated collision energy acquisitions. The resolution and mass measurement accuracy (<5 ppm RMS) obtained from the Q-ToF is key in providing additional specificity for the analysis of complex tryptic digests, allowing confident protein identification.

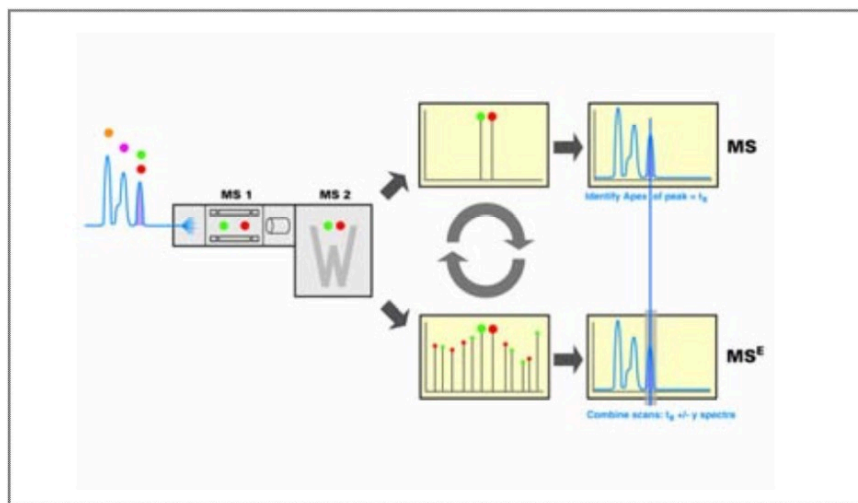
Experimental

The starting point for the Protein Expression System is reproducible sample preparation. The key elements comprise protein solubilization, using the novel surfactant *Rapigest* SF, followed by reduction, alkylation, and subsequently enzymatic digestion using sequencing grade trypsin. Samples of digested protein are subsequently analyzed with highly reproducible online HPLC or UPLC separations with an electrospray Q-ToF mass spectrometer.

During the HPLC run, as peptides elute from the column, the Q-ToF instrument is switched alternately at onesecond intervals between low and elevated collision energy with argon in the hexapole collision cell.

The MS spectra obtained at different collision energies are stored in separate functions. During data acquisition the quadrupole, MS^1 , is not mass selective but rather operates in the radio frequency (rf) only mode. Thus, the quadrupole operates as an rf ion guide and passes all ions to the hexapole gas cell. The first data function (MS) at low energy shows only the normal pseudomolecular ions, while the second at elevated energy (MS^E) shows their associated fragments. At no point during the experiment are precursor ions selected using the quadrupole as in the traditional product ion MS/MS acquisition.

As the mass spectrometer is continually alternating between low and elevated energy on the gas cell, and all ions are passed to the oa-ToF for mass measurement. This results in exact mass measured fragment ions, that can potentially be observed for every peptide precursor ion present in the low energy ToF dataset. Therefore, the two acquired data functions contain the entire set of exact mass measured precursor and product ions that can be formed by fragmentation within the collision cell and represent a 'global expression dataset'.



Schematic showing the principle of electrospray LC-MS data acquisition with the Protein Expression System.

Data Processing - Bioinformatics

The acquired global expression datasets are processed using an algorithm that employs a maximum likelihood technique. This determines the exact mass, intensity, retention time and estimates precision for each eluting peptide species from an LC-MS dataset. Peak areas can then be normalized to an

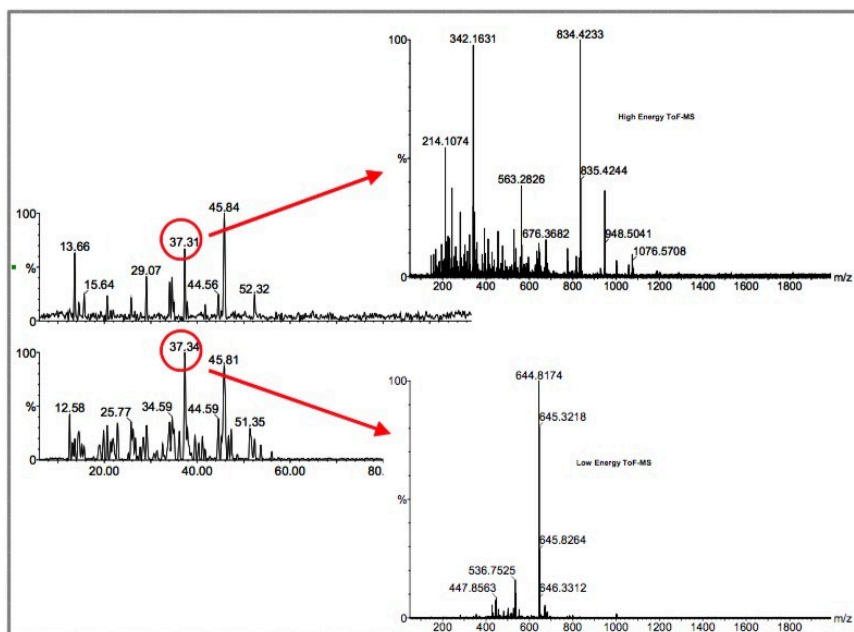
endogenous or exogenous internal standard. This provides the first output which can then be used for relative quantification, as exact mass retention time (EMRT) signatures are generated for every eluting species detected. Output of these EMRT signatures to programs such as Spotfire allows data visualization, or statistical treatment/clustering of the datasets to be performed.

Further bioinformatic processing identifies the apex of each peptide ion identified in the low energy data function, and interrogates the MS^E data function, acquired with an elevated collision energy, to reveal an associated set of exact mass fragment ions. By combining the exact masses of the peptide molecular ions and the associated fragment ions, a highly specific set of time resolved masses are generated that define that peptide. These time resolved sets of exact masses are searched against a databank using a proprietary peptide fragmentation model. Each matching protein in the databank is given a probability and confidence value.

The results are displayed in an interactive browser and list the matched protein sequences ranked by their probability. This display also allows visualization of the associated peptides that match to each protein sequence and display the relevant fragment ion data.

The expression ratio of proteins having a significant probability can then be determined across two or more sample sets, by comparing peak intensities of the relevant monoisotopic peptide masses. The probabilistic technique also provides a measure for the uncertainty of the ratios. Batch processing can be used to compare control with experimental samples.

These routines can be incorporated, as part of the Protein Expression System, in ProteinLynx Global SERVER v2.2.



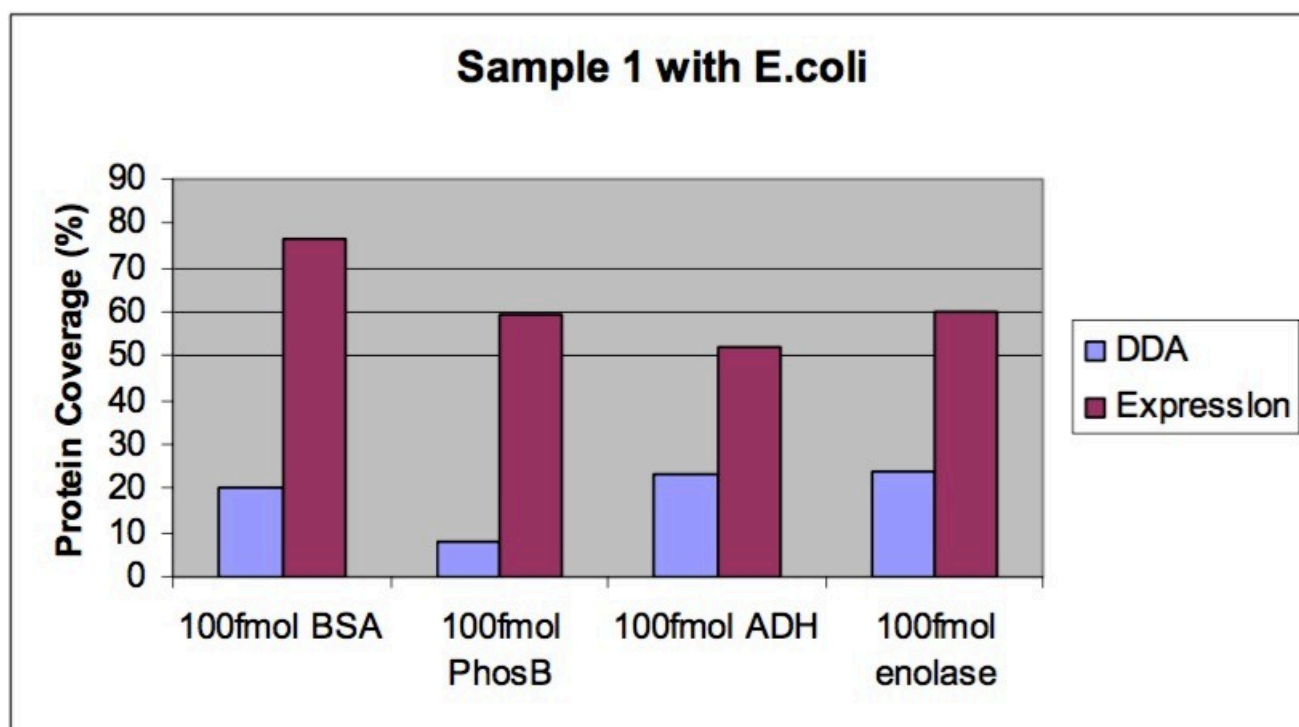
Data acquisition showing the low energy (molecular ions) and elevated energy (fragment ions) at a specific chromatographic time point.

Results and Discussion

The power of the Protein Expression System is shown in the following examples.

Qualitative protein analysis

A sample set containing tryptic digests of five standard proteins of varying molecular weight were used. Samples were analyzed as a mixture of the five proteins and also in the presence of an *E.coli* cytosolic cell fraction tryptic digest. The parallel expression profiling approach was compared to the traditional LC-MS/MS DDA approach.



Results showing the increased sequence coverage obtained for the expression analysis on four proteins compared to a DDA analysis.

Expression profiling of the sample containing all five digested proteins (equimolar concentrations) spiked into the digested *E.coli* cytosolic fraction resulted in positive identification of the standard proteins with high confidence and coverage. A key factor is the excellent mass measurement accuracy that is obtained on the tryptic peptide that match to the protein sequences as shown below.

Protein	RMS error
Serum albumin P02769	2.93 ppm
Enolase P00924	5.08 ppm
Glycogen phosphorylase B P00489	3.24 ppm
Alcohol dehydrogenase P00330	3.40 ppm
Hemoglobin beta chain P02070	2.69 ppm
Hemoglobin alpha chain P01966	1.26 ppm

Mass measurement accuracies obtained from the qualitative identification of proteins from a fivecomponent mixture (Protein Expression System Standards).

Further bioinformatic processing of a single expression dataset, resulted in the identification of over 100 constituent proteins from *E.coli*.

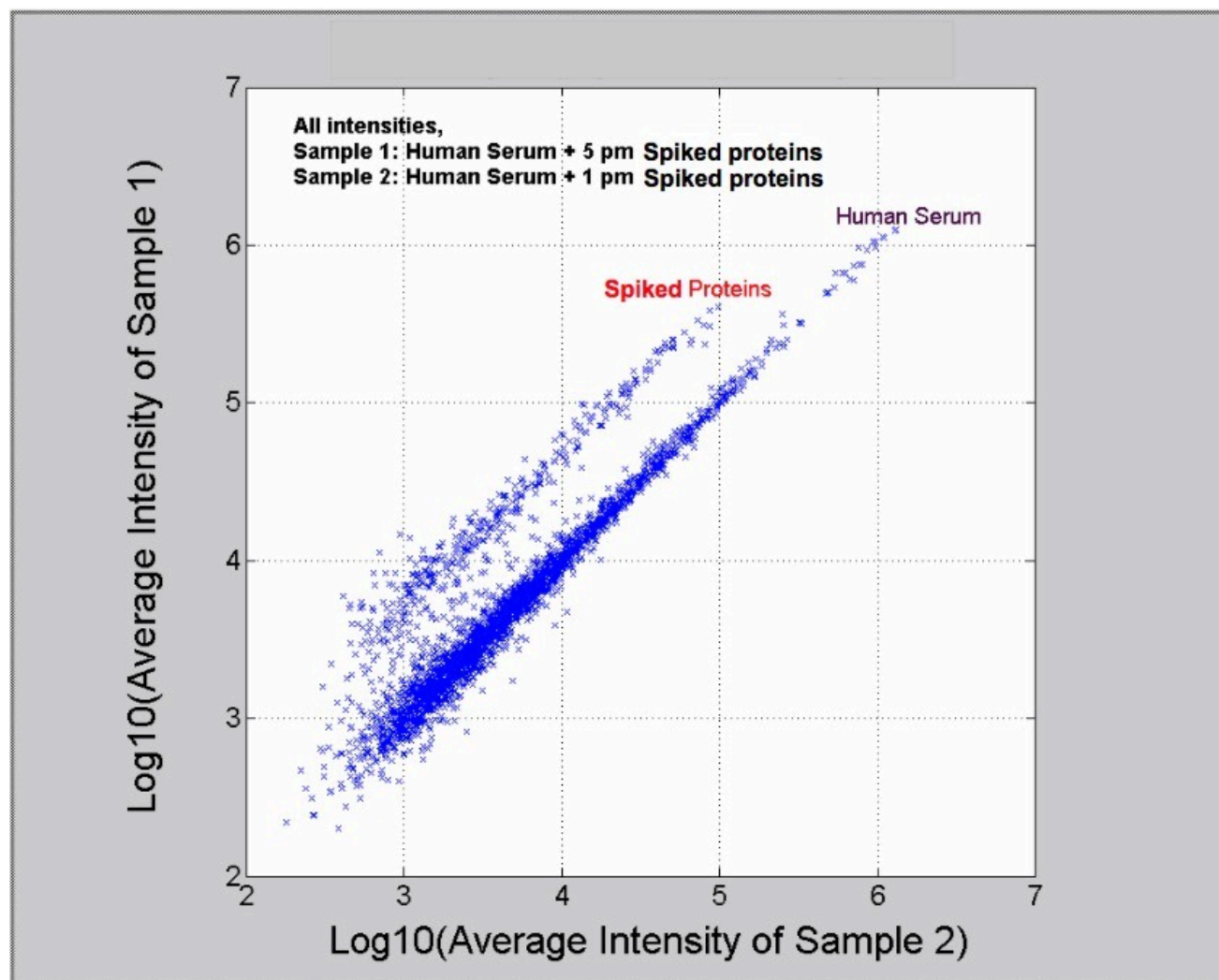
In comparison, the DDA approach identified the standard proteins, but with significantly lower coverage than the results from the Protein Expression System experiment. While from the *E.coli* sample, LC-MS/MS was unable to identify a comparable number of proteins from a single analysis.

Relative quantification

The ability to accurately determine EMRT signatures in each individual sample and subsequently normalize them (using a common endogenous or exogenous set of peptides), allows accurate and reproducible relative quantification to be obtained. This approach does not require the use of stable isotope labelling and is not limited to peptides containing a specific residue (e.g. cysteine in an ICAT experiment). Therefore, sample processing is simplified, with no protein or peptide derivatization required, and protein coverage is increased, with potentially all tryptic peptides from a protein observed in the expression experiment.

The following example shows relative quantification of five standard proteins, at varying

concentrations, spiked into human serum. Comparison of the data extracted from the spiked human serum injections was performed using the Protein Expression System Informatics Software.



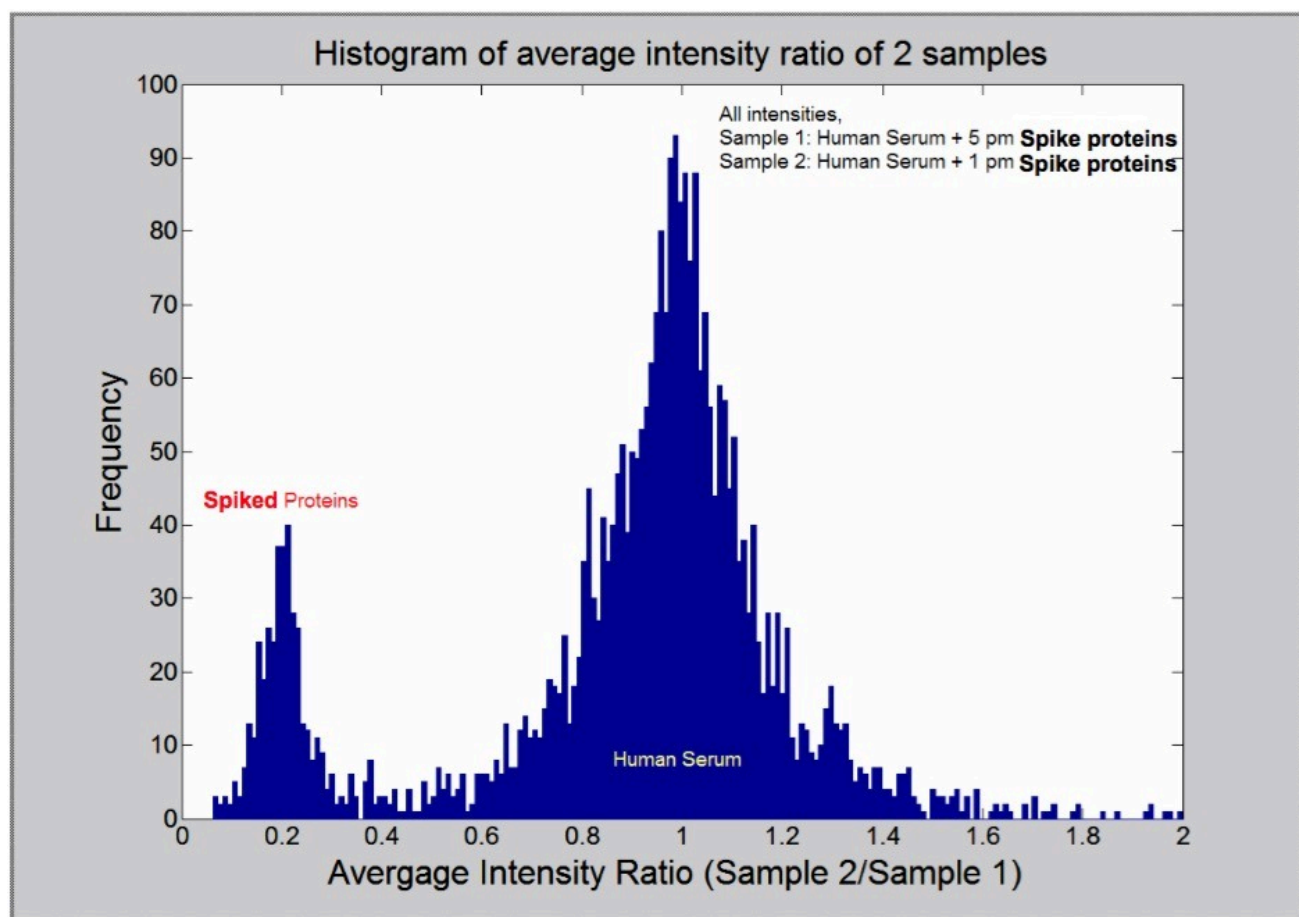
Relative quantification, at the peptide level, of five spiked standard proteins present at two different concentrations in a complex human serum sample.

The log ratio of the EMRT components from two samples in human serum (5.0 and 1.0 picomoles of spiked exogenous proteins) shows the power of this analytical approach.

The cloud of ions lying along the 45-degree axis (see figure on left) represents EMRT peptides that were measured with an intensity ratio of approximately 1.0. These peptides are associated with

human serum peptides. The employed methodology allowed these ions to be measured over nearly 4 orders of magnitude in intensity. The cloud of ions to the upper left section of the human serum proteins illustrate the EMRT signatures associated with the tryptic peptides of the spiked exogenous protein digests.

This intensity ratio information can also be further displayed as a frequency plot (see figure below). The intensity of the two EMRT groupings clearly illustrate the 1:1 ratio of the human serum peptides and the 1:5 ratio of the spiked exogenous tryptic peptides.



Frequency plot of the EMRT signatures obtained from the comparison of two samples of human serum spiked with two levels of five standard proteins. A differential expression ratio of 1:5 is clearly observed.

This approach of comparing EMRT signatures across sample sets allows the monitoring of protein

expression changes with high accuracy. This is evidenced by the coefficient of variation of the intensity measurements made on the human serum peptides, where variations of generally less than 20% are observed.

Conclusion

As demonstrated, the Waters Protein Expression System provides a paradigm shift in proteomic capabilities. The utilization of novel chemistry products in combination with excellent chromatographic reproducibility, exact mass MS data from the Q-ToF and proprietary bioinformatics allows high quality qualitative proteomics. Due to the high reproducibility of the data, it is possible to also perform relative protein quantification on the same datasets. This is a global approach that is ideal for the study of protein expression and the discovery and characterization of biomarkers.

Featured Products

ACQUITY UPLC M-Class System <<https://www.waters.com/134776759>>

ProteinLynx Global SERVER (PLGS) <<https://www.waters.com/513821>>

720000910, June 2004

© 2022 Waters Corporation. All Rights Reserved.
