Improving Productivity with Applied Biosystems GPS Explorer™ Software

Purpose

GPS Explorer™ Software is the application layer software for the Applied Biosystems 4700 Proteomics Discovery System. The software is intended to significantly improve the productivity of proteomic researchers involved in identification and quantitation of proteins from gel spots or LC fractions through intelligent job creation and confident data interpretation and analysis. This product bulletin details the functions and workflows using GPS Explorer™ Software.

Overview

Confident protein identification using tandem mass spectrometry is one of the most important aspects of protein analysis. Unmatched spectral quality, accurate mass measurement, and information-rich data are all key features of the 4700 Proteomics Discovery System. GPS Explorer Software takes full advantage of this high-quality spectral data by utilizing results-driven features for both protein identification and expression analysis.

Key Features

- Results-driven multi-tiered acquisition for both in-gel digestion and multi-dimensional liquid chromatography (MDLC) workflows
- Expression dependent experiments using ICAT® reagents for targeted analysis of differentially expressed proteins i.e., MS/MS identification of proteins that are changing
- Intuitive graphical user interface for easy setup and visualization of intelligent acquisition routines, data interpretation and results, with a highly flexible custom report generator
- Remote bi-directional access to set up, control, and monitor acquisition, processing, and results from the 4700 Proteomics Discovery System
- Access to protein molecular functions, biological processes, and comprehensive genomic information through direct links to the Celera Discovery System™ *

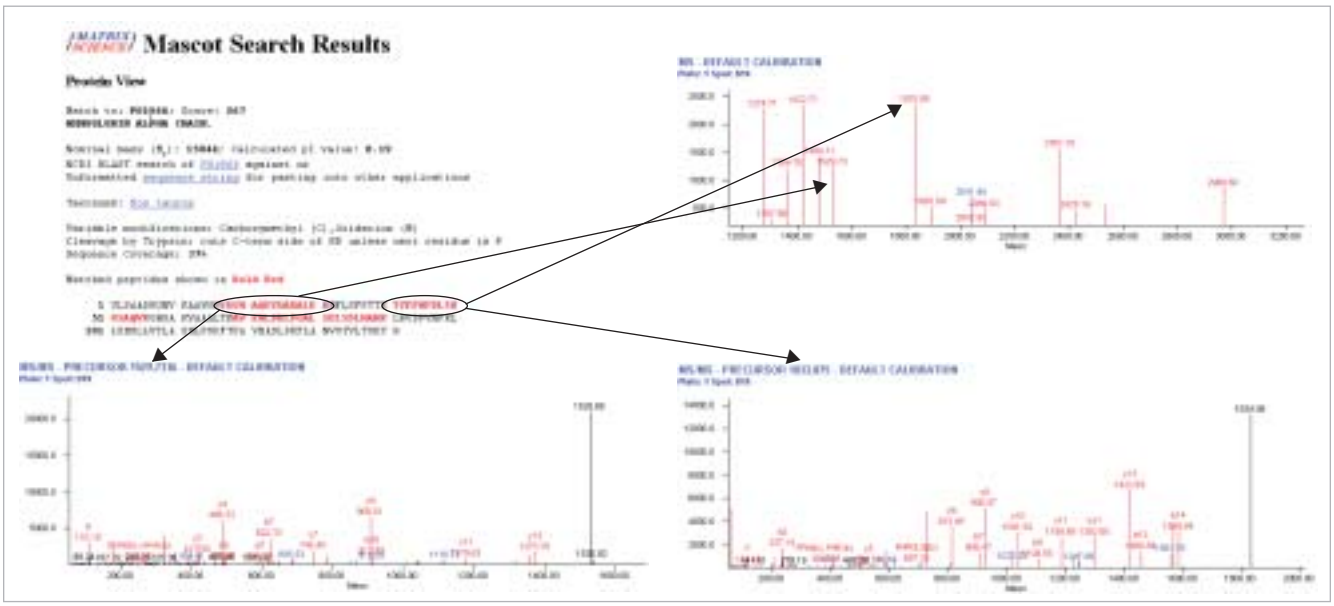

Figure 1. In the PMF confirmation mode, peptides that match the top protein hit in MS mode are automatically selected for MS/MS analysis to reliably confirm protein identification.

Workflows

Sample preparation for protein identification by mass spectrometry typically revolves around two workflows. The most common approach is to perform in-gel digestion on samples prepared from 1D or 2D gels followed by protein identification by peptide mass fingerprinting (PMF). A second approach, the LC MALDI workflow, uses multi-dimensional liquid chromatography (MDLC) and microfraction collection to separate and collect samples directly on the MALDI target. The LC MALDI-based workflow can provide greater depth of coverage and more reliable protein identifications in complex mixtures. GPS Explorer Software is uniquely designed to accommodate both the in-gel digestion and the LC MALDI approach.

Gel-based Workflow

For samples analyzed from 1D or 2D gels where one MALDI spot represents one gel spot, users can set up data acquisition and search routines that involve automatic PMF analysis followed by intelligent selection of precursors for MS/MS based on PMF search results. This MS/MS data is then used for a second round of database searching for protein identification. Two modes of analysis based on first pass PMF search results are supported in the software.

- PMF confirmation mode—All the proteins identified by the PMF search are confirmed by performing MS/MS on a user-specified number of peptides matching each protein (Figure 1). This increases the overall confidence of protein hits.

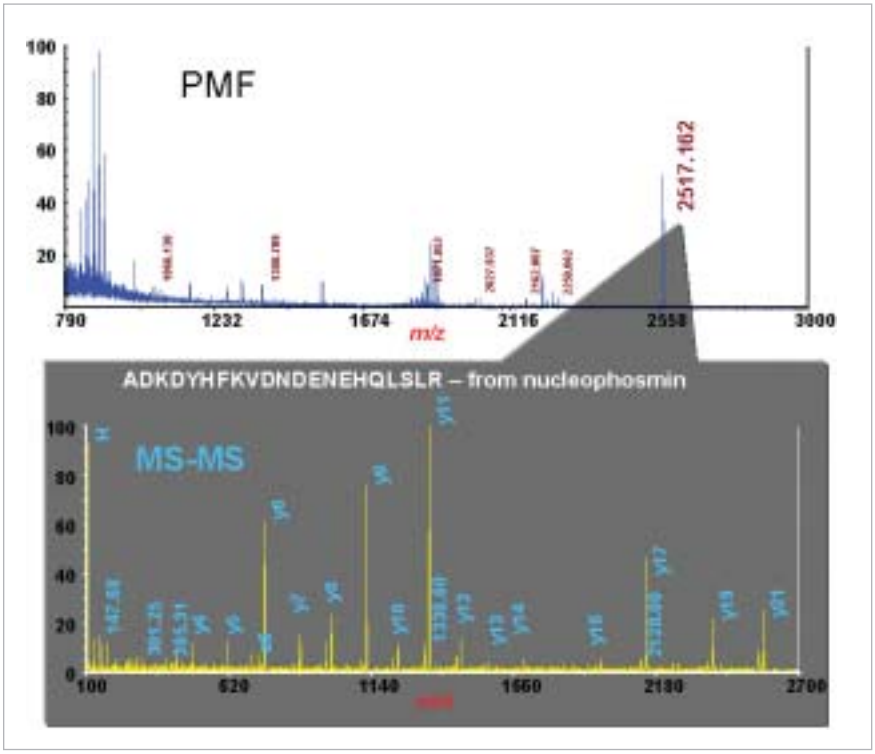

Figure 2. In the PMF dig-down mode, additional proteins are identified that would otherwise be missed. In this example, the PMF contained only three peaks from the major protein component, not yielding a significant match. Only MS/MS data could identify this protein unambiguously.

- PMF dig-down mode—Unassigned peptides from the PMF search are automatically sent for subsequent MS/MS analysis for additional protein identifications (Figure 2). This key advantage of the 4700 Proteomics Discovery System and GPS Explorer™ Software allows the identification of additional proteins in the sample that otherwise would be missed.

LC MALDI Workflow

Using the LC MALDI workflow, complex protein mixtures can be directly deposited onto a MALDI plate, for example, by using an LC Packings PROBOT™ Micro Fraction Collector, creating high-resolution chromatography "frozen-in-time" on the surface of the sample plate. High quality MS and MS/MS data can be generated on large numbers of peptides from the protein mixture, greatly improving overall proteome coverage. Precursor ions can be selected from the spot containing the highest concentration of that peptide. In addition, and in contrast to on-line chromatography techniques, the LC MALDI approach reduces ion suppression effects, thereby increasing the overall dynamic range observed—orders of magnitude typically.

- Proteome-wide ID—Automated MS data acquisition is followed by MS/MS acquisition and combined database searching to identify large numbers of proteins in a mixture. Peptides that are identified in each spot are then mapped to proteins that are identified across the entire plate.
  - Without the time constraints of on-line chromatography, the instrument can perform MS/MS on as many precursors as desired.

- For peptides that span multiple spots, precursor ions are selected from the spot with the highest concentration (Figure 3).

- The graphical interface allows users to visually see where the peptides map for a given protein across the sample plate (Figure 4).

Expression Dependent Analysis

The LC MALDI workflow is the perfect complement to expression dependent analysis of complex mixtures using ICAT® Reagents. Quantitation of the mass tagged pairs is performed to determine expression levels. The software then automatically normalizes the pair ratios and selects only the truly differentially expressed peptides above the specified threshold on which to perform MS/MS analysis and identification. In a typical sample, only 10-20% of all the proteins are differentially expressed. Automated expression dependent analysis focuses valuable MS/MS time on the proteins that are changing (Figure 5). Alternatively, because the user has control over the ratio threshold for invoking MS/MS, all peptide pairs can be fragmented by setting the threshold appropriately.

Innovative Architecture

GPS Explorer™ Software allows bi-directional connection to multiple instruments and remote access to results through an integrated Oracle® database. Automated multi-tiered results-based analysis allows flexibility in workflows. Large sample sets of data are imported from the 4700 Proteomics Discovery System and submitted to the searching algorithm

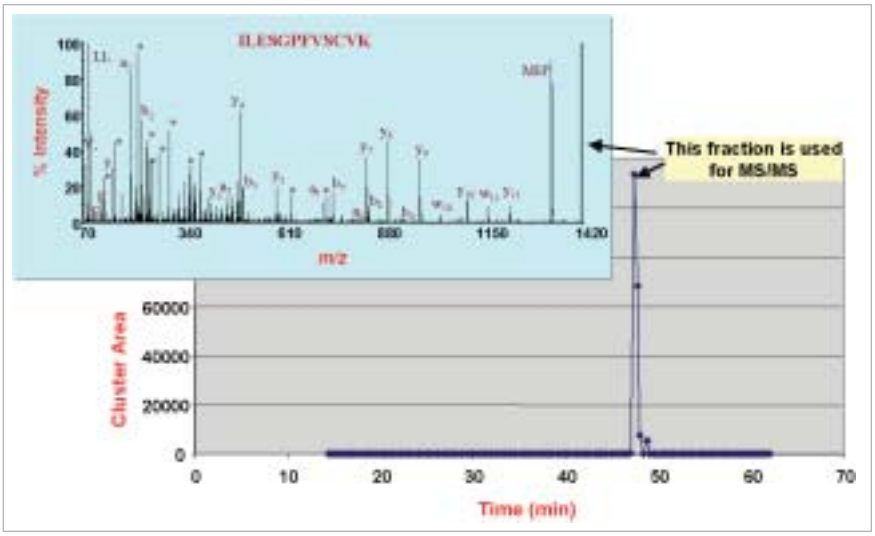

Figure 3. When peptides span multiple HPLC fractions, GPS Explorer™ Software intelligently selects the spot with the highest concentration for further MS/MS analysis. In the above example, the peptide with sequence 'ILESGPFVSCVK' elutes over four spots as depicted by the blue dots on the extracted ion chromatogram. The spot with the highest concentration is chosen for MS/MS. The inset shows the MS/MS data obtained for this peptide.

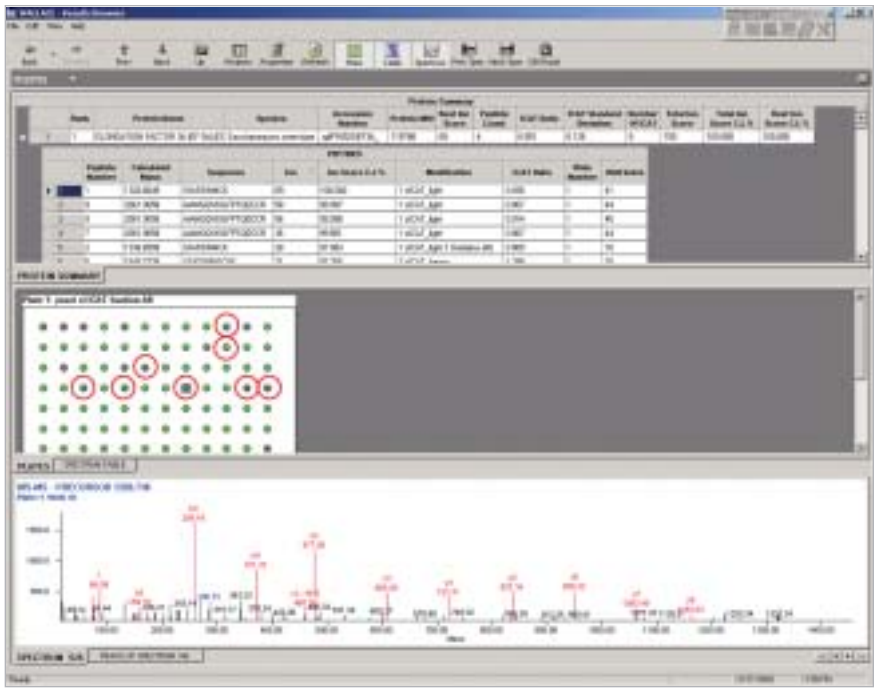

Figure 4. In the Results Browser, peptides that were identified and their corresponding proteins are listed in the top pane (the protein *Elongation Factor 3A* is expanded to show the peptides that were identified). The plate view in the center pane depicts where the peptides map for a given protein across the sample plate (wells circled in red correspond to peptides identified for *Elongation Factor 3A*). The bottom pane displays the MS/MS data for the peptide selected from the top pane—in this case, the peptide from well 41 with sequence SAVIIDNMCK.

as MS only data, MS/MS data, or as a combined search of MS and MS/MS to provide the most complete coverage of spectral information for protein identification.

Viewing Results and Report Creation

An intelligently designed graphical user interface allows easy interpretation and visualization of matched peptide spectra from acquired MS and MS/MS data. Convenient color-coding of spectral peaks quickly differentiates matched from unmatched peaks in the database search. Additionally, the MALDI plate viewer displays proteins identified with varying degrees of confidence in different colors (Figure 6). Confident identification is quickly determined using exclusive scoring algorithms based on MASCOT® database scoring. The MASCOT database probability scoring scheme is based on the size of the database searched. The significance score threshold—the score used to discriminate good hits from others—varies depending upon the number of peptides searched, which in turn depends upon the modifications used, mass tolerances, etc. Scores are simplified to provide better methods of comparing peptide and protein data from different search parameters and databases. Custom reports can be generated in a variety of formats and exported in xls, pdf, rtf, txt, and htm file formats. All results are stored in a back-end relational database for easy and flexible access.

Integration with the Celera Discovery System™ (CDS)* and Other Programs

Once a protein is identified, the next step is to determine the biological significance of that protein. What is the function of that protein? What other proteins are similar? What are the interacting protein partners? To help answer these questions and more, Applied Biosystems now offers direct links to the Celera Discovery System through GPS Explorer™ Software. With the critical biological information found in CDS, users can more intelligently plan their next research experiment.

- Information about molecular functions and biological processes for a protein are displayed within the GPS Explorer Software Results Browser.

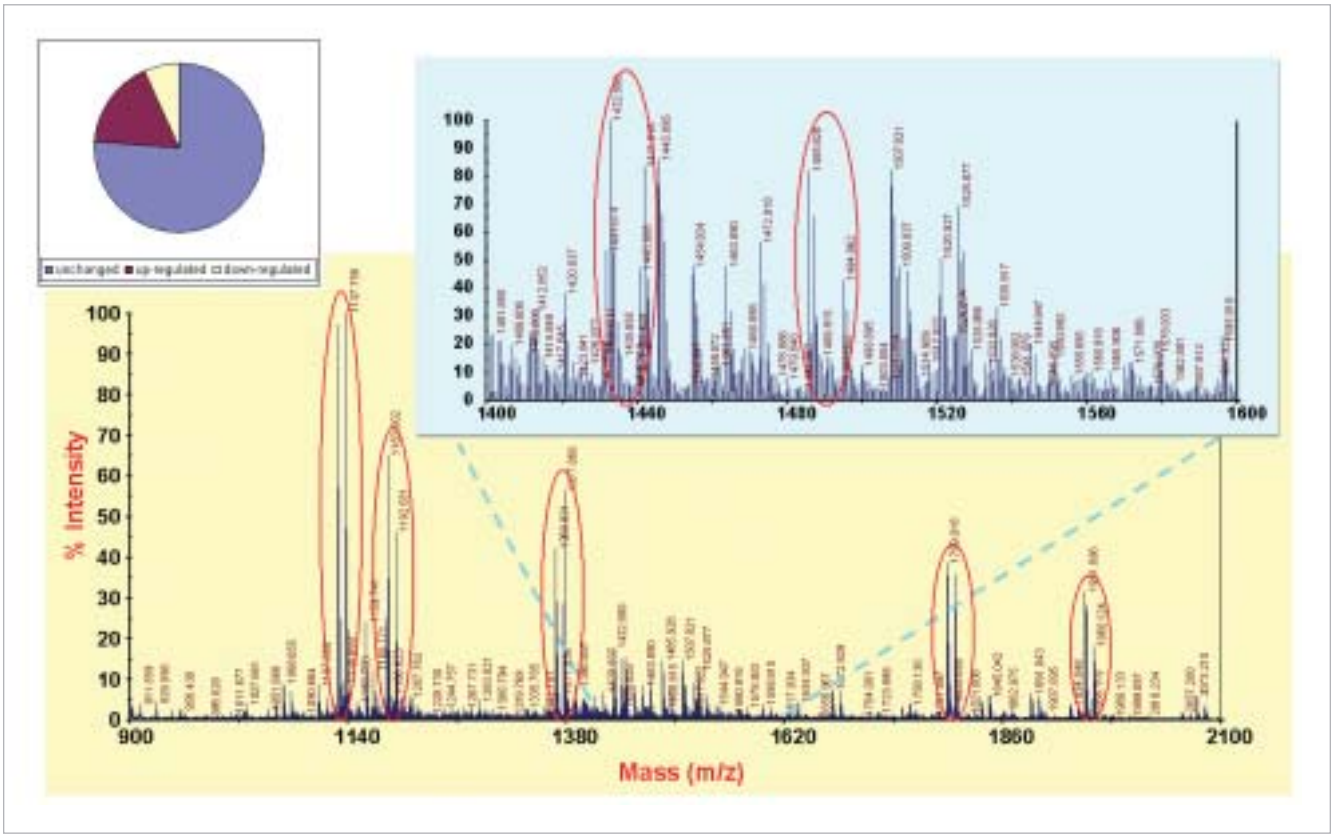

Figure 5. In a differential expression experiment only the most highly expressed proteins are automatically selected for MS/MS analysis, thereby saving time and focusing analysis on proteins of interest. To demonstrate the complexity of a typical experiment, this example of TGF stimulated lung data contains more than 200 ICAT® reagent labeled peptide pairs identified in a single HPLC fraction. Some selected ICAT reagent pairs are circled in red. Overall, over 1,000 ICAT reagent pairs were quantified with most of those unchanged in their expression levels (~75%). Because the ratio threshold for MS/MS is set to only fragment differentially expressed proteins, valuable instrument time is not wasted on identifying proteins that are not changing.

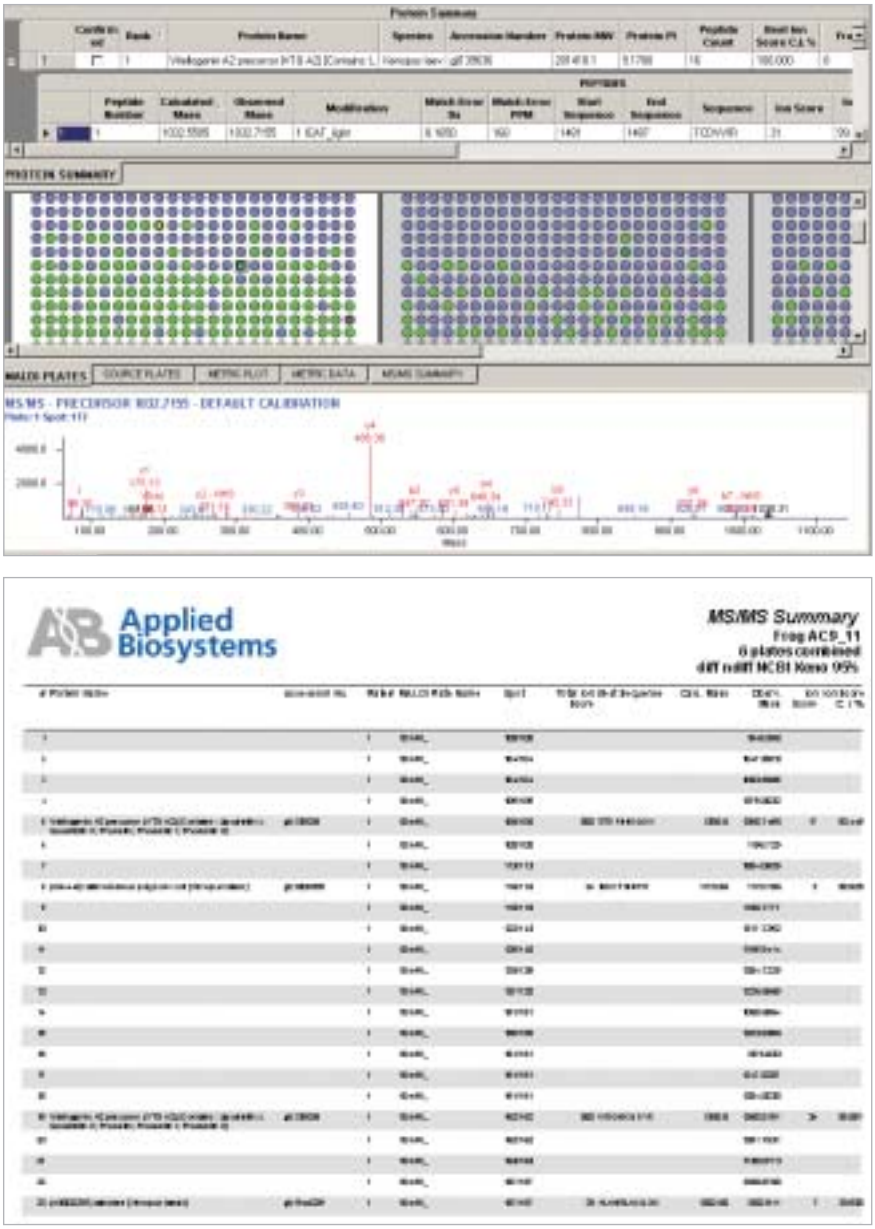

Figure 6. Color-coded graphical user interface for results display and custom report generation make GPS Explorer™ Software extremely user friendly.

- Direct links to the Celera Discovery System™ enable detailed biological analysis using Panther ontologies. Additional information on chromosome locations, gene structures, motifs, domains and sequence similarities between genomes for that protein is also available (Figure 7).

Once classification information has been assigned to a set of proteins, other programs such as Spotfire DecisionSite™ for Functional Genomics can be used for the exploration and comparison of protein expression patterns and their relationship to protein function. Proteins with similar expression patterns can be clustered, correlated with gene ontology information, compared with corresponding gene expression levels, and correlated across multiple tissue types.

Conclusions

GPS Explorer™ Software is a comprehensive data analysis package that significantly improves productivity by supporting both traditional gel- and LC MALDI-based proteomics workflows for qualitative and quantitative analysis. Multi-tiered automated results-based analysis routines with feedback to data acquisition provide powerful experimental strategies for getting the most information out of your samples. For example, the PMF confirmation and dig-down modes allow identification and confirmation of proteins that might otherwise get missed. Expression dependent analysis of ICAT® Reagent prepared samples focuses MS/MS identification experiments on only those proteins that are changing. Additionally, graphical results visualization tools make manual inspection straightforward, while flexible report generation allows results to be presented in a custom format. Finally, the integrated Celera Discovery System provides researchers easy and direct access to fundamental biological information on the proteins identified. In short, GPS Explorer Software provides an intelligent package of software tools for getting the most comprehensive information from your samples in the shortest time.

** Celera Discovery System requires a subscription for access. To learn more about CDS subscriptions and pricing, contact your sales representative or visit the Celera Discovery System Web site at: http://www.celeradiscoverysystem.com/signup/home.cfm.*

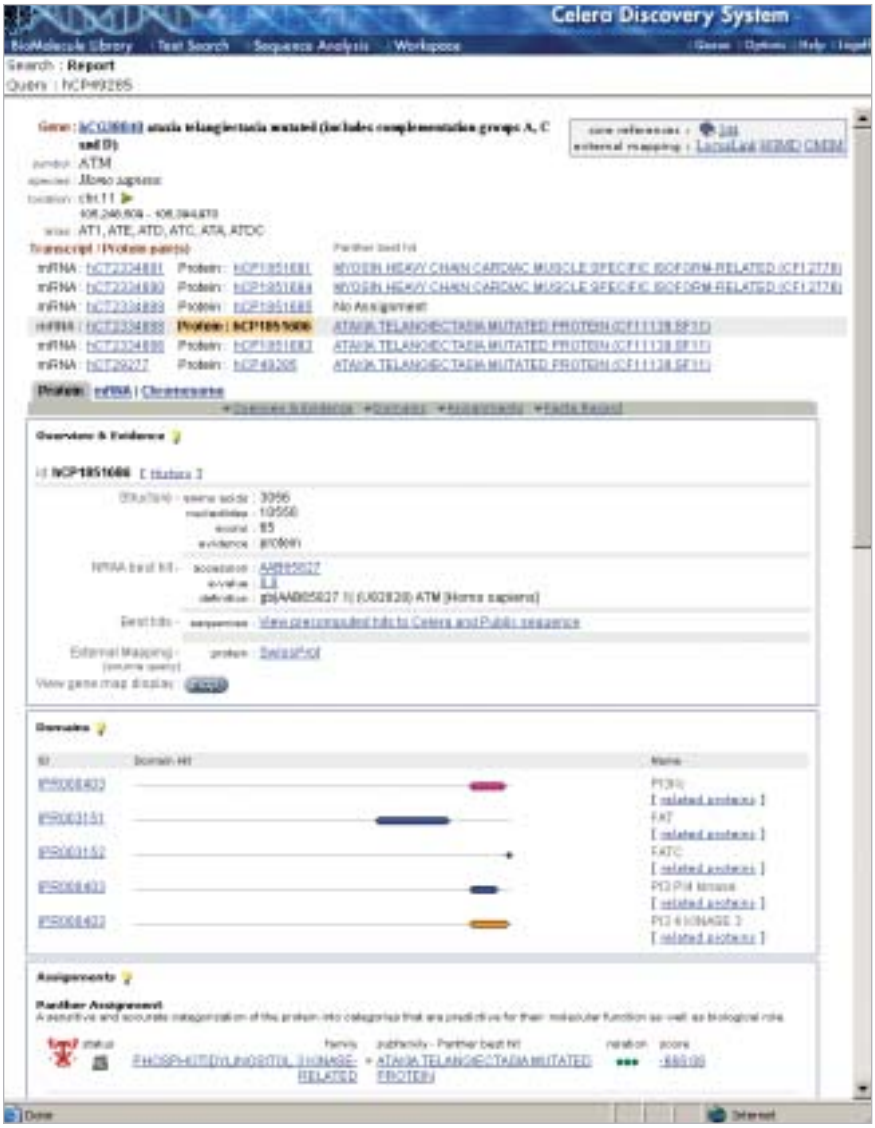

Figure 7. Direct links to the Celera Discovery System™ (CDS) provide additional information after a protein is identified.

iScience.

iScience. To better understand the complex interaction of biological systems, life scientists are developing revolutionary approaches to discovery that unite technology, informatics, and traditional laboratory research. In partnership with our customers, Applied Biosystems provides the innovative products, services and knowledge resources that make this new, **Integrated Science** possible.

Worldwide Sales Offices
Applied Biosystems vast distribution and service network, composed of highly trained support and applications personnel, reaches 150 countries on six continents. For international office locations, please call the division headquarters or refer to our Web site at www.appliedbiosystems.com.

Applera is committed to providing the world's leading technology and information for life scientists. Applera Corporation consists of the Applied Biosystems and Celera Genomics businesses.

Headquarters
850 Lincoln Centre Drive
Foster City, CA 94404 USA
Phone: 650.638.5800
Toll Free: 800.345.5224
Fax: 650.638.5884

For Research Use Only.
Not for use in diagnostic procedures.

Printed in the USA, 08/2003
Publication 115PB04-01

Applied Biosystems