

# Automated Decoy Analysis in Proteomics Projects

Peter Hufnagel<sup>1</sup>;  
Ulrike Schweiger-Hufnagel<sup>1</sup>;  
Wolfgang Jabs<sup>2</sup>; Markus Lubeck<sup>1</sup>;  
Gerhard Koerting<sup>2</sup>; Ray Sanchez<sup>2</sup>;  
and Detlev Suckau<sup>1</sup>

<sup>1</sup> Bruker Daltonik GmbH, Bremen/ Leipzig, Germany  
<sup>2</sup> Protagen AG, Dortmund, Germany  
<sup>3</sup> Bruker Daltonics, Billerica, USA

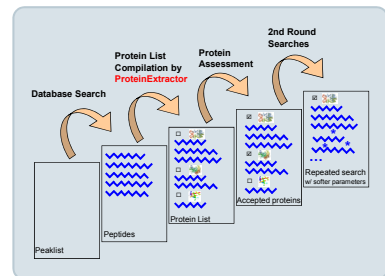
## Introduction

Result integration and validation are key issues for the identification and quantification of proteins on great numbers. For a maximized number of identified proteins, one strategy comprises biological and technical replicates, another involves separation steps on protein and / or peptide level. The resulting redundant search results need to be integrated on peptide level.

The principle of artificial Decoy protein sequences is a widely accepted answer to the task of automated result validation. Nevertheless, the actual realization of the Decoy approach varies among research groups and search engines. Here, we apply a composite database of real and Decoy protein entries.

**Fig. 1: Database searches in ProteinScope.**

1. Search engines report identified peptides.
2. The ProteinExtractor combines the peptides into a non-redundant protein list.
3. The proteins are assessed following a Decoy strategy for a minimized False Positive Rate.
4. Selected proteins can be submitted to a 2nd Round search for the detection of PTMs etc.



## Methods

### Sample

- > 200 µg of human lung carcinoma cell lysates (line A549; TGF β treated)
- > 21 bands from a 1D SDS PAGE gel
- > Tryptic digest

### LC ESI-MS/MS

The peptides were separated on a nano-LC system and supplied to an ion trap mass spectrometer (HCTUltra, Bruker).

### ProteinExtractor and Decoy

ProteinScope (Bruker) performed database searches (Fig. 1) using a composite Decoy IPI human database on a Mascot search engine on the data of each gel band separately. The 21 peptide lists were combined by the ProteinExtractor algorithm (part of ProteinScope), which generated one integrated protein list. All proteins below a False Positive Rate of 5 % were automatically accepted.

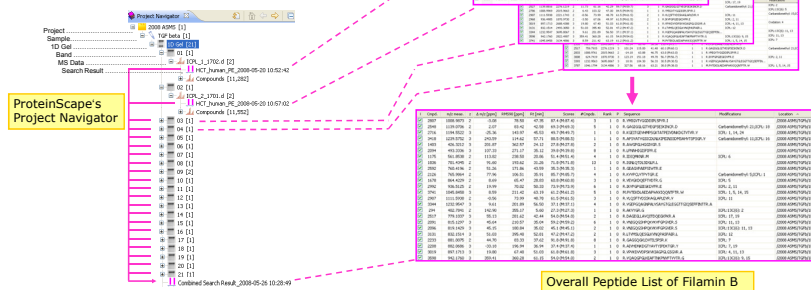
## Results

### ProteinExtractor

894 proteins and 5472 peptides resulted from the combination of the 21 gel band results.

Details for one protein are shown in Fig. 2, right. Filamin B has been identified from four bands, each with 15 to 25 peptides. ProteinExtractor has created a common non-redundant peptide list for Filamin B with 38 peptides. 35 were accepted in the automatic assessment step.

The fact that Filamin B peptides are found in various bands shows that the protein separation power of the gel was not sufficient. This is often observed when subjecting complex proteomes to SDS PAGE alone.



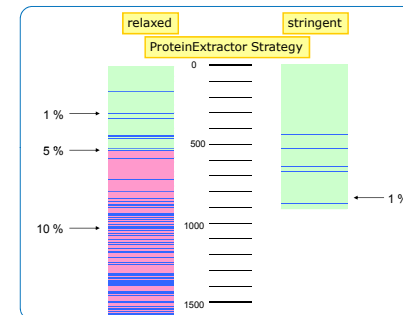
**Fig. 2: Data handling in ProteinScope.**

The Project Navigator of ProteinScope (left) displays the 1D gel workflow incl. LC-MS/MS data and search results which have been compiled by the Protein Extractor

## Decoy

The parameters of the Protein Extractor are typically set in a **stringent** way. This means, that the result list contains only proteins with at least two reliably identified peptides (Fig. 3, right). For demonstrating the Decoy strategy, we also applied a more **relaxed** setup with a softer score threshold (Fig. 3 left). This resulted in a longer protein list, however it contained more Decoy proteins.

## ASMS 2008, WPX-638



**Fig. 3: ProteinExtractor setups.** More relaxed score criteria lead to a longer protein list with more False Positives.

The blue lines mark the artificial protein entries in the composite database. The green and the red area mark the ranges with a False Positive Rate below and above 5 %.

## Conclusions

The **ProteinExtractor** compiles a non-redundant protein list from peptide lists of different origin.

Days of manual processing time are condensed into hours of computing time.

The **Decoy Strategy** validates the number of identified proteins according to a desired False Positive rate.

- (1) into one respective protein list for each band and
  - (2) into one integrated protein list for the whole gel.
- On the right, the peptide lists for Filamin B for bands 1 to 4 and the overall peptide list are shown.