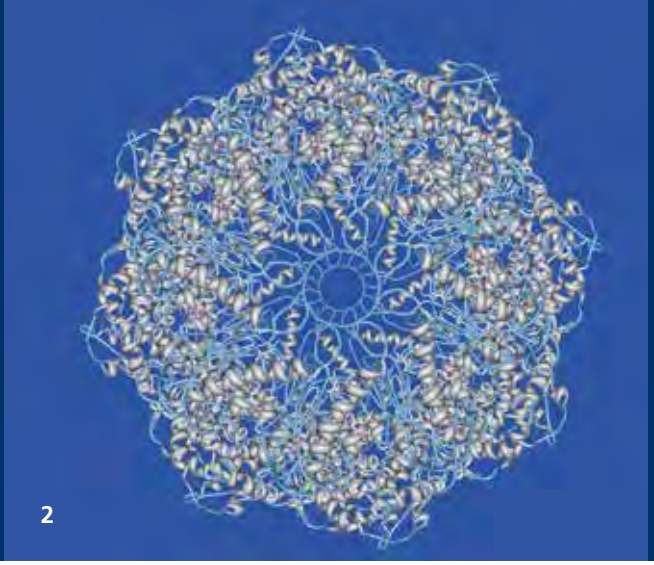
Fraunhofer
IGB

FRAUNHOFER INSTITUTE FOR INTERFACIAL ENGINEERING AND BIOTECHNOLOGY

PROTEIN AND PROTEOME ANALYTICS
IDENTIFICATION, CHARACTERIZATION AND EXPRESSION ANALYSIS OF PROTEINS



2

PROTEINS – PROTEOMICS

Proteins, macromolecules made up of amino acids, are major building blocks of all cells. They shape the cells structure and as molecular "machines" are responsible, for example, for the transport of substances, the catalysis of chemical reactions and the recognition of signaling molecules. Amino acids, which are linked by peptide bonds to form chains of up to several thousand units, serve as building blocks. In this context, the sequence of amino acids determines the structure and function of the protein.

The totality of all proteins in an organism, a tissue, a cell or a cell compartment under exactly defined conditions and at a specific point in time is termed proteome. As a result of continuous de novo synthesis and the concurrent degradation of proteins, its composition changes continuously. These changes are influenced by environmental stimuli and controlled by complex regulatory processes. Depending on the species, the proteome can comprise up to 1,000,000 protein species and thus normally exceeds the number of gene sequences encoded in the genome many times over. This is primarily the result of mRNA splicing and subsequent (posttranslational) modifications of the proteins.

Posttranslational protein modifications occur subsequent to translation. These changes are usually initiated by the organism or by the cells themselves and can be influenced by environmental factors. Examples of the multifarious possibilities are phosphorylation, which is mediated by protein kinases, hydroxylations of proline residues or glycosylations. Some of these processes occur at the protein synthesis site, others occur in certain organelles or also outside the cell.

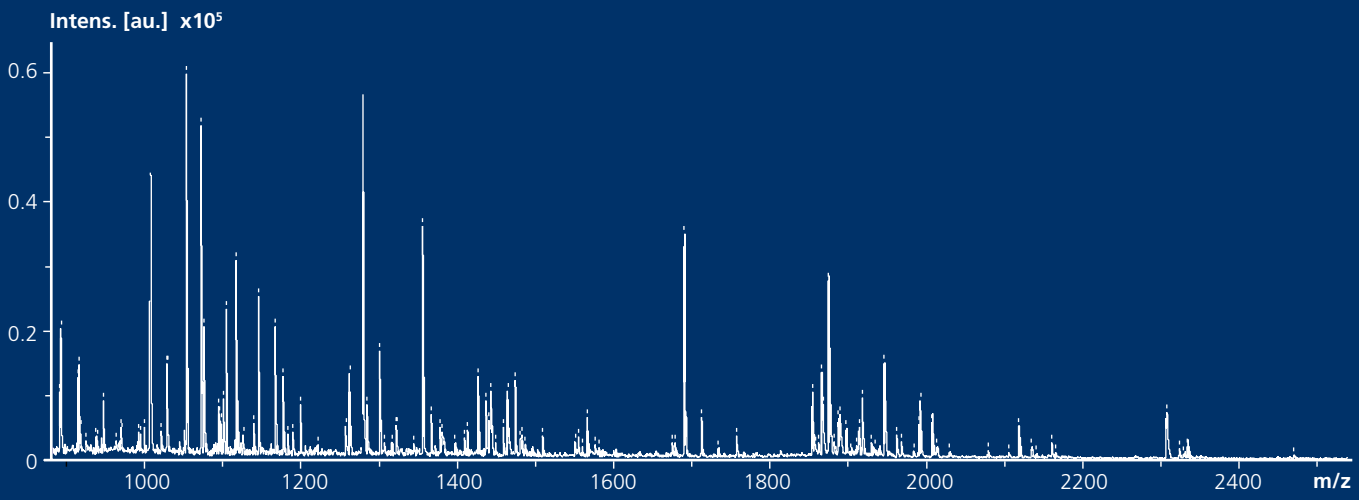
Applications

In the past decade, the importance of mass spectrometry for protein analyses has greatly increased. It can be used both to identify a protein and to analyze the amino acid sequence of a so far unknown protein. In the process, possible alterations in the side chains can be detected, such as those occurring in posttranslational modifications. In combination with different sample preparation procedures, mass spectrometry is an important tool in the repertoire of current methods for analyzing individual proteins or entire proteomes.

The techniques used for protein analytics at Fraunhofer IGB comprise gel electrophoretic, chromatographic and mass spectrometric methods, which can be used for a broad application spectrum:

- Protein identification
- Protein characterization – posttranslational modifications
- Expression analysis – proteomics
- Identification and validation of biomarkers
- Quality control of expressed proteins

1 *Gel electrophoresis under UV light.*
© BioRegio STERN/Lichtenscheidt.
2 *3D depiction of a chaperonin protein.*



1

PROTEIN IDENTIFICATION

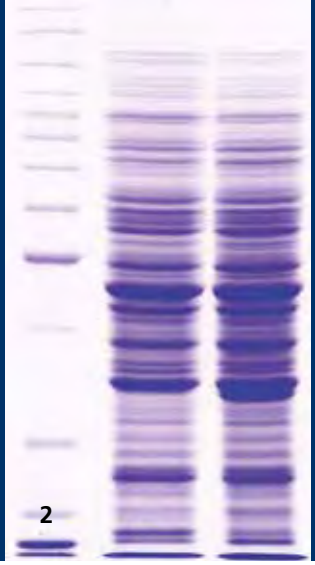
With the aid of mass spectrometry very small quantities of one or several different proteins can be identified. An option for the identification of very small quantities of proteins is their analysis by means of a so-called peptide mass fingerprint (PFM). This is performed subsequent to cleavage of the proteins by sequence-specific proteases. By comparing the experimentally determined masses of the peptides with the protease cleavage of theoretically generated fragment masses of known proteins in a database, conclusions can be made as to the identity of the protein. Because of the existing, extensive genome sequences of a large number of organisms, this method is frequently successful. Additionally, MS/MS analyses of selected peptides for the determination of PSD fragments (post source decay) can be conducted. The degradation pattern of the peptides measured in this manner contains sequence information and can be used to identify them. As a result, potential peptides (masses), which could not be assigned with the previously conducted PMF, can also be investigated, e.g. in the case of identification of a protein from a gel band or a gel spot (SDS-PAGE, 2D-PAGE). In this manner, additional proteins can be identified. If high protein sequence coverage is required, the peptide mixture can be separated in analytical scale before the mass spectrometric analysis by means of *reversed-phase* chromatography. This enables the detection of a larger number of peptides with MS and to analyze them by means of MS/MS.

An additional approach for the identification of a protein does not require any previous proteolytic cleavage. By means of the mass spectrometric analysis of the ISD fragments (in source decay) of the intact protein, the amino acid sequence can be read from both N and the C terminus and then used for the identification.

Application examples

Bottom-up protein identification

Latex contact allergies are triggered by rubber tree (*Hevea brasiliensis*) proteins, which are found in latex products such as disposable gloves or catheters. At Fraunhofer IGB we have analyzed the protein composition of the different fractions which arise in the latex production process including the products (medical gloves) by means of one- or two-dimensional polyacrylamide gel electrophoresis (1D-, 2D-PAGE) in order to identify allergenic latex proteins. Allergenic proteins are identified with using blood sera of people allergic to latex. The identity of these proteins was subsequently determined by means of PMF and MS/MS analyses. For identification, a database with 10,829 EST sequences (expressed dequence tag) of *Hevea brasiliensis* (extracted from the National Centre for Biotechnology Information NCBI) was established on the MASCOT server. In total, we were able to identify 15 proteins, five of which had not yet been described previously in connection with latex allergies.

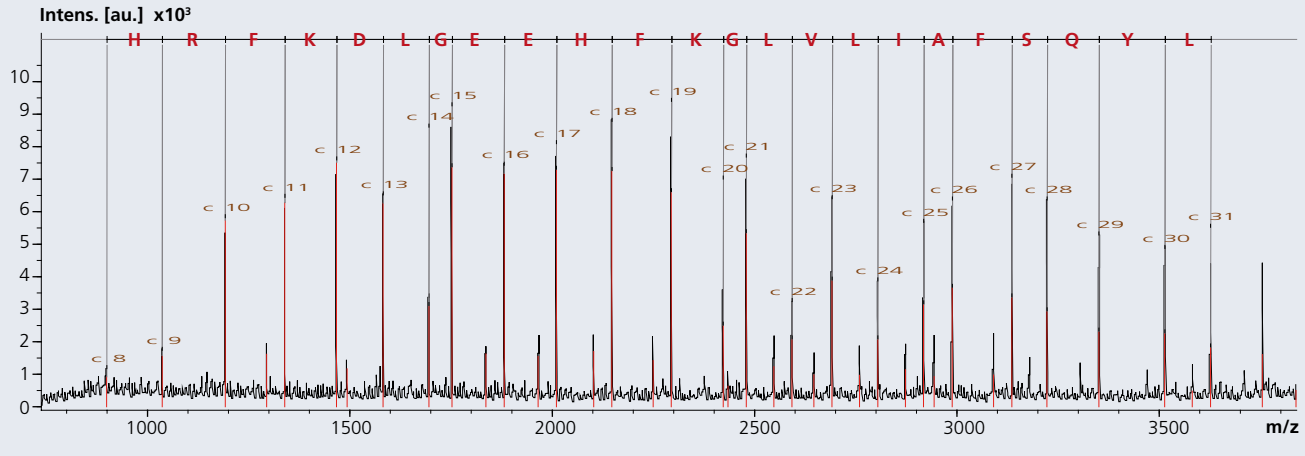


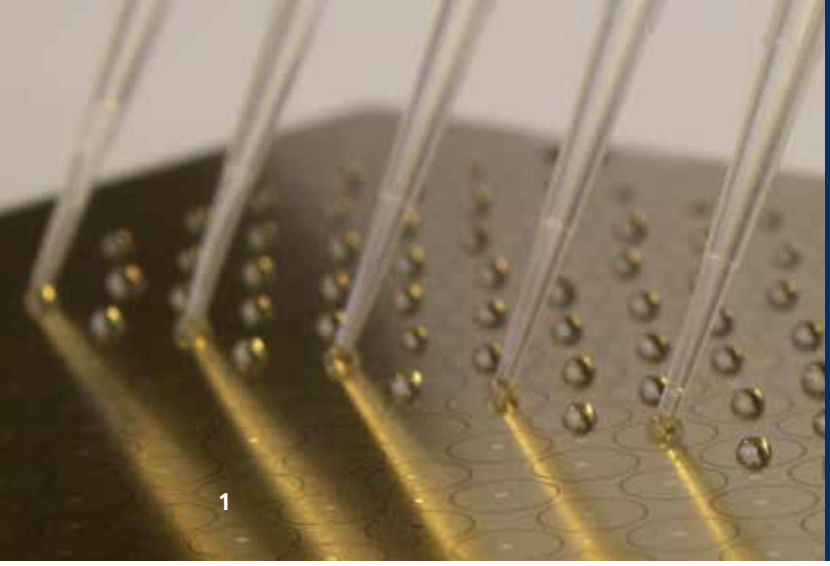
Top-down protein identification

In the scope of a study organized by the *Protein Sequencing Research Group* of the *Association of Biomolecular Resource Facilities* in 2009, two samples, which each contained one protein that was not described in detail, were sent to the participants. They were to determine the identity of the proteins and as large a region of respective N-terminal amino acid sequence as possible. For this purpose, a mass spectrometric analysis of the ISD fragments (in source decay) of these proteins was performed at the Fraunhofer IGB. With its help, the identification of the N-terminal sequence of up to 32 amino acids and the proteins as well as the tags remaining from the expression could be identified. Furthermore, part of the C-terminal sequence could be determined. The identity of the proteins was confirmed by peptide mass fingerprint analysis and subsequent MS/MS investigations.

1 *MS spectrum of a peptide mass fingerprint.*
2 *Separation of a protein mixture using SDS-PAGE.*
3 *Natural latex is obtained from the sap of rubber trees.*

Annotated in-source decay MS spectrum.





PROTEIN CHARACTERIZATION — POSTTRANSLATIONAL MODIFICATIONS

An initial characterization of a protein can be performed by the acquisition of different parameters. Statements can be made on the protein's relative content of acidic and basic amino acids by means of isoelectric focussing. The quantitative determination of the amino acids which make up a protein can be performed by means of chromatographic amino acid analysis in the central Analytics Service Unit at the Fraunhofer IGB.

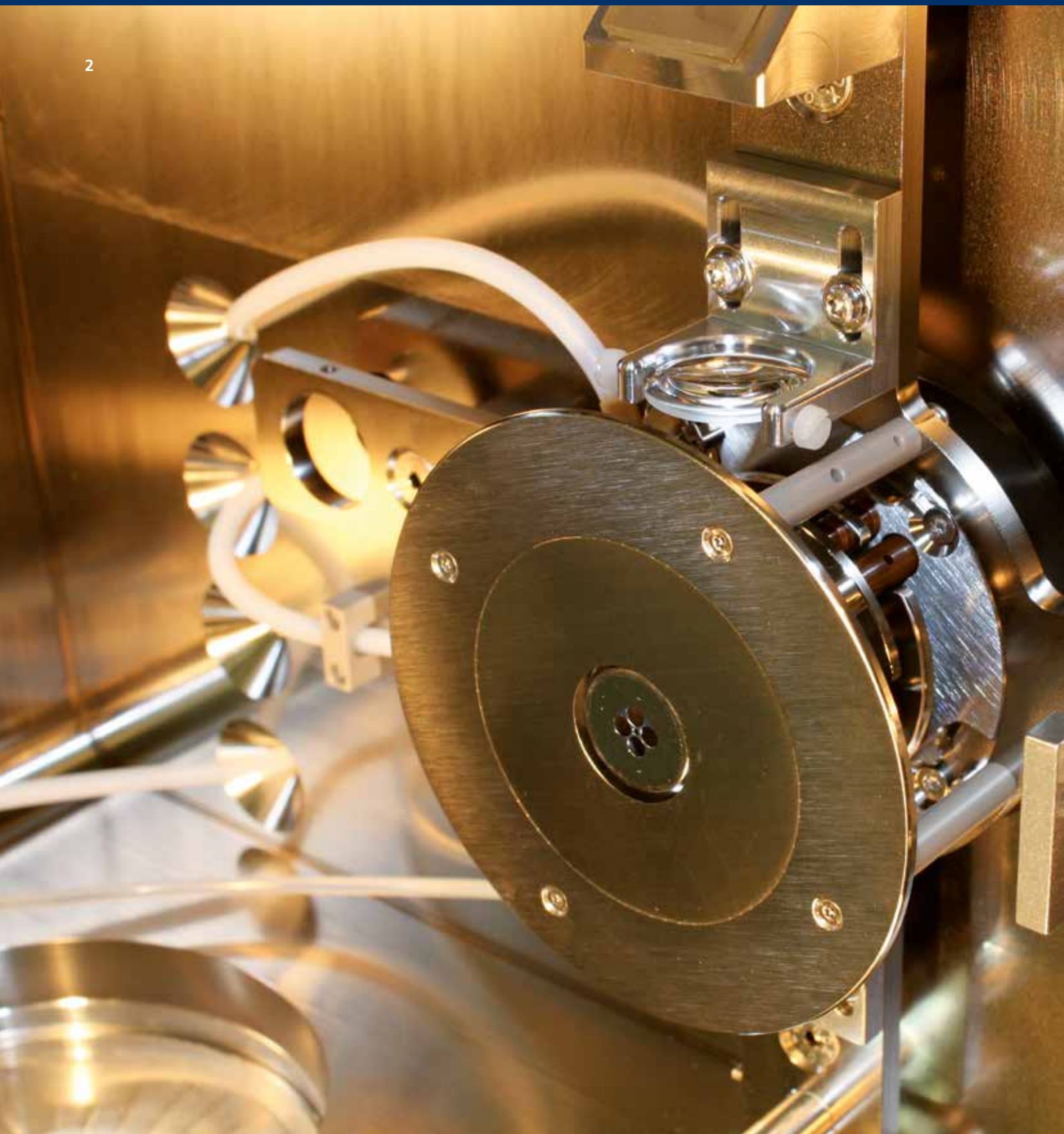
Beyond this, the use of mass spectrometry allows collection of information on the exact mass, the amino acid sequence and possibly existing (posttranslational) modifications. By means of the mass spectrometric analysis of the ISD fragments of a protein, it is possible to read the amino acid sequence from both the N and the C terminus. If the protein is cleaved by one or more different proteases, the peptides produced can be analyzed and also sequenced by LC-MALDI and MS/MS such that sequence information on additional regions of a protein can be obtained. The sequence data can also be used for the investigation of posttranslational modifications, which can be determined from changes in the mass data.

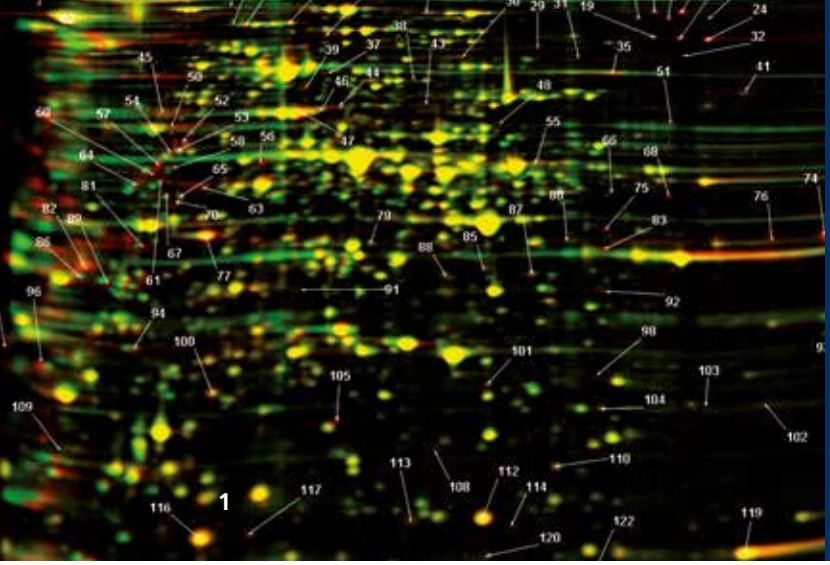
Application example

Characterization of the matrix protein collagen

Collagen, one of the most frequent proteins in the mammalian proteome, is of great importance for medicine and biotechnology. Thus, for example, it is used as a natural three-dimensional matrix for mammalian cells in tissue engineering. Collagen, which could be obtained in near native state by means of an alternative isolation technique, was characterized by mass spectrometry at Fraunhofer IGB. The mass spectrometric characterization of the thus-purified collagen showed that it exhibits a 2.3 kDa higher mass than customarily isolated collagen. Using MS/MS analysis, we were able to trace this back to a reduced degradation of the collagen due to the improved purification procedure. Moreover, posttranslational modifications were identified, which provided new insights into the structure of collagen *in vivo*.

1 *MALDI sample carrier.*
2 *MALDI ion source.*





1

2

EXPRESSION ANALYSIS – PROTEOMICS

The proteome of an organism is subject to continuous changes in its composition. This high degree of complexity was the special challenge of this investigation. To reduce its complexity, a proteome is normally separated into a number of fractions. At Fraunhofer IGB two-dimensional gel electrophoresis (2D-PAGE) is used to separate protein extracts according to their isoelectric point and their molecular mass; and fluorescent staining is subsequently employed to detect and quantify them. The determined frequency of a protein in different samples is compared using statistical procedures. In this manner, the protein's expression behaviour can be monitored under different conditions. The proteins whose expression level differs under the selected conditions are subsequently identified by means of mass spectrometry.

Application example

Pathogenic mechanisms of *Candida albicans*

Candida albicans is a facultative pathogenic yeast, which can cause a broad spectrum of diseases particularly in immunocompromised patients. They range from superficial mucosal infections up to systemic candidiasis. To identify the proteins which are important for the pathogen's infection potential, differential proteome analyses with 2D-PAGE were performed at Fraunhofer IGB. In these investigations, the infection was simulated by *C. albicans* on human epithelial tissues *in vitro*. The differential protein expression analysis showed a definite regulation of proteins which belong to different metabolic pathways. The identification of the proteins from the gel spots was performed subsequent to tryptic digestion by analyzing the resulting peptides by means of peptide mass fingerprinting. These identifications were verified by subsequent MS/MS analyses. The suitability of these proteins as biomarkers is currently being studied.

The proteins secreted by *Candida albicans* participate in the interaction of the fungus with the host and can, for example, also serve as biomarkers for candidiasis. In order to identify the proteins released by *C. albicans*, they are isolated from the culture medium and subjected to tryptic digestion. The resulting peptide mixture was separated by reversed-phase chromatography and deposited on a MALDI sample carrier (LC-MALDI) in 192 fractions. We were able to identify 30 proteins by means of a subsequent MS/MS analysis of circa 400 peptides.

3

SEARCHING FOR BIOMARKERS

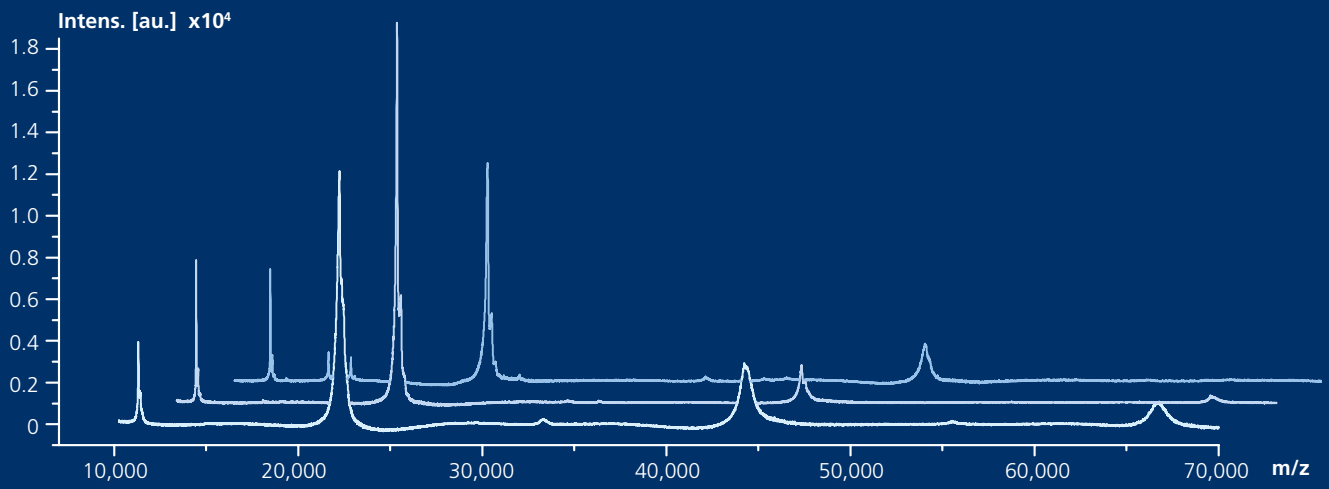
Processes occurring in the body are frequently distinguished by characteristic biological attributes, so-called biomarkers. They can be, for example, peptides or proteins, which only occur in a complex sample in case of a disease and thus are characteristic for it. New diagnostic tests can be developed on the basis of these biomarkers.

In this context, the challenge in the search for new biomarkers is to identiy them among the multitude of molecules which are present in a blood or tissue sample. Normally, this is achieved by a chromatographic separation of the sample components and a subsequent comparison of different samples. By using special separation techniques, e.g. reversed-phase chromatography with nano- or microliters flow rates on an analytical scale and the highly sensitive mass spectrometry as a detection technique, it is possible to also acquire the low molecular weight components (≤ 10,000 Da) of complex samples. The subsequently applied statistical procedure of principal component analysis allows comparison of different samples and the identification of characteristic components for a special condition.

Two-dimensional gel electrophoresis can be used for separation of proteins as well. In combination with graphic evaluation software this allows the comparison of different, complex samples. This technology is also capable of depicting different posttranslational modifications. Mass spectrometry is then subsequently used to identify a protein which can be used as biomarker from the gel.

If the spatial distribution of a protein or other compounds in a tissue is of interest, mass spectrometry can also be used. The imaging MALDI-MS or MALDI imaging mass spectrometry allows the visualization of two-dimensional distribution of proteins or other substances on thin sections of a sample. The distribution image of an investigated mass generated by this procedure can also be superimposed with light-microscopic images of histological stains and thus, for example, be used to investigate potential biomarkers.

1 *Separation of a protein mixture using 2D gel electrophoresis*
2 *Candida albicans.*
3 *Nano-LC with connected LC-MALDI fraction collector.*



1

QUALITY CONTROL OF EXPRESSED PROTEINS

Special requirements are required for proteins prepared for medical use. Thus, among other things, it must be possible to verify the exact amino acid sequence of the relevant protein. By means of mass spectrometric analysis, it is possible to verify the amino acid sequence from both the N and the C termini (ISD). After appropriate sample preparations, which can also include a directed cleavage of the protein by proteases and the chromatographic separation of the resulting products, it is possible to verify the correct amino acid sequence with high sequence coverage (LC-MALDI-MS/MS). In this analysis, contingently present posttranslational modifications can also be investigated.

Application example

Modification of pharmaceutical proteins

Biopharmaceutically employed active substances or diagnostics can be conjugated using chemical methods with polyethylene glycol (PEG). In this manner a coating is applied, which results in a reduced immunogenicity, an elevated stability against proteases and a retarded excretion. Thus, the active substance can develop its action more efficiently. At Fraunhofer IGB mass spectrometry with MALDI-TOF/TOF was used to provide proof of a successful conjugation with appropriately modified proteins and to characterize them.

1 *MS protein spectra.*
2 *MALDI-TOF/TOF mass spectrometer.*

Technical equipment

Mass spectrometer
- Ultraflex II MALDI-TOF/TOF
- Finnigan LCQ^{DECA} Mass spectrometer

HPLC
- Ultimate 3000 nano-LC
- Surveyor LC System

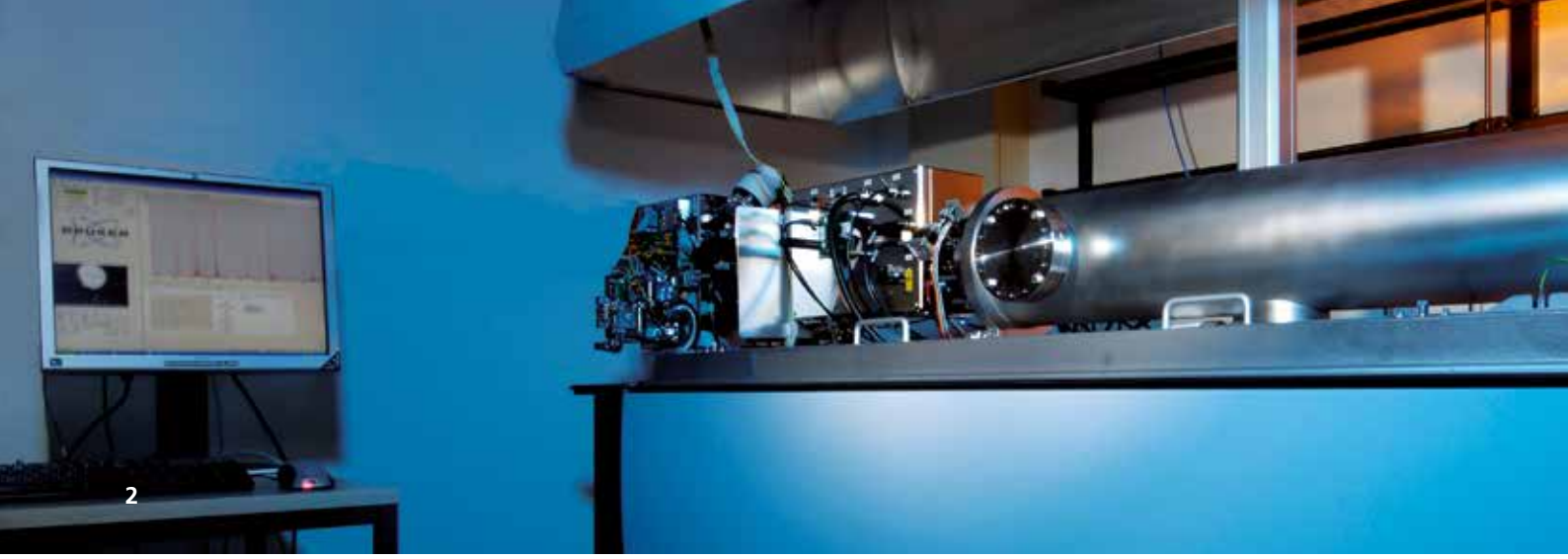
LC-MALDI Fraction collector
- Proteineer fc

Isoelectric Focusing
- Ettan IPGphor II

2-D Electrophoresis
- Protean® plus DodecaTM cell
- Protean II xi Cell

Laser gel scanner
- Image Reader FLA-5000 Series

Software
- Bruker Compass
- MASCOT Server
- Xcalibur
- Proteome Discoverer
- Delta2D
- Advanced Image Data Analyzer (Aida)



RANGE OF SERVICES

The Fraunhofer IGB provides know how around the mass spectrometry platform. In this context, our focuses are centered around the field of identification and characterization of proteins as well as expression analysis. Both, complete analysis of individual samples as well as individual steps during protein analysis will be provided according to the customers demands.

Protein identification and characterization

- Purification or concentration of the proteins by means of precipitation, filtration or chromatographic methods
- Identification of the proteins via ISD analysis
- One- or two-dimensional polyacrylamide gel electrophoresis with subsequent sensitive staining of the proteins
- In-gel digestion of selected bands or spots with proteases or chemical agents
- Purification and concentration of the peptides
- Reversed-phase chromatographic separation of peptide mixtures with automatic transfer of the fractions to mass spectrometric sample carriers
- Selection/set-up of an appropriate MASCOT database for the identification
- Identification of the proteins via peptide mass fingerprint analysis
- Identification of the proteins by means of MS/MS (PSD or CID analysis)
- *De novo* sequencing of peptides with the aid of chemical modification to improve the quality
- Specific purification of modified proteins or peptides
- Identification of posttranslational modifications in the MS/MS or ISD spectra

Expression analysis and search for biomarkers

- Sample generation including quality control
- Purification or concentration of the proteins by means of precipitation, filtration or chromatographic methods
- One- or two-dimensional polyacrylamide gel electrophoresis with subsequent sensitive staining of the proteins
- Scanning of the gels with subsequent evaluation and statistical analysis to detect differentially expressed proteins
- Identification of the proteins
- Reversed-phase chromatographic separation of peptide mixtures with automatic transfer of the fractions to mass spectrometric sample carriers
- Sensitive detection of the ingredients by means of MALDI-MS
- Comparative analyses of different samples by means of a principal component analysis to determine the differences
- Identification of the proteins by means of MS/MS (PSD or CID analysis)

Together with the existing microarray facility, the classical chemical analytics and the fermentation technology at Fraunhofer IGB, system biological approaches can also be followed. The focuses in this context are on the cultivation of microorganisms and eukaryotic cells in bioreactors as well as in the quantitative analytics of nucleic acids, proteins and metabolites.

Fraunhofer Institute
for Interfacial Engineering
and Biotechnology IGB
Nobelstrasse 12
70569 Stuttgart
Germany

Phone +49 711 970-4401
Fax +49 711 970-4200
info@igb.fraunhofer.de
www.igb.fraunhofer.de

CONTACT

Dr. Anke Burger-Kentischer
Phone +49 711 970-4023
anke.burger-kentischer@igb.fraunhofer.de

apl. Prof. Dr. Steffen Rupp
Phone +49 711 970-4045
steffen.rupp@igb.fraunhofer.de

Fraunhofer IGB brief profile

The Fraunhofer Institute for Interfacial Engineering and Biotechnology IGB develops and optimizes processes and products in the fields of health, chemistry and process industry, as well as environment and energy. We combine the highest scientific standards with professional know-how in our competence areas – always with a view to economic efficiency and sustainability. Our strengths are offering complete solutions from the laboratory to the pilot scale. Customers also benefit from the cooperation between our five R&D departments in Stuttgart and the institute branches located in Leuna and Straubing. The constructive interplay of the various disciplines at our institute opens up new approaches in areas such as medical engineering, nanotechnology, industrial biotechnology, and environmental technology. Fraunhofer IGB is one of 69 institutes and independent research units of the Fraunhofer-Gesellschaft, Europe's leading organization for applied research.

www.igb.fraunhofer.de

1803 – BR mbt – protein-proteomanalytik – en | printed on paper from responsible sources | © Fraunhofer IGB