netherlands proteomics centre

highlights

npc highlights 1 | 2005

Featuring cutting-edge research projects of the Netherlands Proteomics Centre

Frontpage

Preparation of samples for Matrix-assisted laser desorption/ionisation-time of flight mass spectrometry (MALDI-TOF MS). The technique is used for highly sensitive detection and characterization of peptides and proteins.

Contents

2 About
3 Preface
4 Highlight
4 Jeroen Krijgsveld, Qualitative and quantitative analysis of large datasets in proteomics
8 Egbert Boekema, Systematic searching for novel complexes of membrane proteins
12 Roland Pieters, Old concepts and new methods detect galectins in complex biological samples
16 John Strouboulis, Characterisation of transcription factor complexes isolated by in vivo biotinylation tagging
20 Sander van der Krol, Cell-specific N-glycoprotein profiling leads to protein markers
24 Ron Heeren, Mass spectrometer identifies and localises biomarkers in cells
28 Perspective On Proteomics: Cindy Bell, V.P. Genome Canada

highlights 1

> About

The Netherlands Proteomics Centre (NPC) is a strategic collaboration of research groups from six universities, three academic medical centres and several biotech companies. With a scientific program addressing key areas of proteomics in 56 projects, and specialised 'research hotels', the NPC performs high-quality research and knowledge transfer in an international context. The NPC is part of the Netherlands Genomics Initiative.

In NPC Highlights researchers present progress and results from NPC projects of the scientific program and the research hotels. NPC Highlights is published by the Netherlands Proteomics Centre.

Netherlands Proteomics Centre
Sorbonnelaan 16
NL 3584 CA Utrecht
t +31 30 253 4564
e info@npc.genomics.nl
w www.netherlandsproteomicscentre.nl

Editorial Board
Albert Heck, Scientific Director NPC
Werner Most, Managing Director NPC
Lubbert Dijkhuizen
Rob Liskamp
Silvie Hermsen, Communications NPC

Coordination and editing
Marian van Opstal
Bèta Communicaties, The Hague

Design
Lude Franke, Ludesign, Utrecht

Photography frontpage
Ivar Pel, Utrecht

To subscribe to NPC Highlights please send an e-mail with your full name, organisation and address to info@npc.genomics.nl

© 2005 Copyright Netherlands Proteomics Centre

preface

Welcome Before you lies the first issue of NPC Highlights, a publication in which we would like to present on a regular basis examples of cutting edge research carried out within our program. NPC Highlights is intended as an information platform for participants in the NPC, as well as for those who have a general interest in proteomics or an interest in how a national program such as the Netherlands Proteomics Centre is developing. We hope you are pleased with this initiative and enjoy the articles herein.

Advanced proteomics research and technology development is at the core of the NPC mission. In the first year, the NPC has succeeded in setting-up a national consortium of researchers from several Dutch universities, academic medical hospitals and industry. With substantial contributions coming from these institutes and companies combined with additional government funding, the NPC was able to make serious investments in scientific personnel and equipment. The centre's research program includes numerous joint projects with academic and industrial research groups via the NPC Research Hotels.

All of the participants in the NPC and many others interested in proteomics came together for the first time at the NPC's grand opening in February 2005. With over 300 attendees, the meeting clearly demonstrated the vitality of the Netherlands' proteomics community and the great interest in sharing views, expertise and knowledge.

The first NPC progress report was recently published (available for those interested). Although we are still in the start-up phase, the progress report clearly demonstrates that the NPC has already surpassed all the milestones that were set for the first year, not only in number of scientific publications but also in number of patents applied for, number of start-up companies initiated as well as in societal outreach activities. After one year we can indeed say that the NPC is now fully up and running.

Over the coming years the NPC will continue to promote proteomics research and stimulate interactions within this field and genomics in general.

A highlight of each of the 6 NPC Research Themes has been selected for this first issue and is presented to you by young NPC researchers.

NPC Executive Board

jeroen krijgsveld

qualitative and quantitative analysis of large datasets in proteomics

Despite the enormous improvements in protein analysis by combing purification techniques and mass spectrometry, quality control of these measurements does not attract much attention. The inclusion of false positive identifications in particular can send research in the wrong direction, thus wasting time and money. Techniques to increase the confidence in protein identification are available. A pledge for implementation of a quality control routine in the analysis of complex mixtures.

In today's proteomics, mass spectrometry has become the method of choice for protein identification. This has largely been driven by technological advances, not only in mass spectrometry, but also in separation techniques such as 2-dimensional electrophoresis and (capillary) chromatography. The maturation of all these techniques has dramatically increased analytical throughput and sensitivity. Identifying hundreds of proteins in a single experiment is now ordinary practice in many labs [1]. In essence, proteomics, as well as many other 'omics' approaches, is a technology used as a discovery tool.

Despite this technological focus in proteomics, its mission has always been to be an integrated part of functional biology. In fact this fit between analytical chemists and practitioners in biology is close to ideal, since one of the driving forces in proteomics has been the question of how to best generate hypotheses, while functional experiments are required for their subsequent biological validation.

With this role of proteomics in a multi-disciplinary setting in mind, this paper will discuss some issues that are at the heart of proteomics technology and that arise when large sets of identified proteins are to be used as a foundation in subsequent biological experiments. An often-overlooked question is how certain these identifications are, and how some routines to improve confidence can be of use. This paper suggests ways of how to perform quality control tests that may help answering these questions.

Additionally, some aspects of protein quantification are discussed. This technique is of considerable interest from a biological point of view. Some recent developments such as full labelling of metabolites are described, as well as improved experimental designs realised in concerted efforts by biologists and mass spectrometrists.

Protein identification Peptides are ideally identified by mass spectrometry by determining their sequence from a given fragmentation spectrum followed by a search through a protein database. However, the average quality of a routinely generated spectrum allows the deduction of a partial protein sequence at best, often in a manual and per-spectrum fashion. Today, it is rather common practice to match uninterpreted spectra directly to a protein database to find the best scoring

What this research is about:

Essentials for top quality experiments and analysis

Mass spectrometry combined with separation techniques is the method of choice for analysis of complex protein mixtures. Hundreds or even thousands of proteins are identified in a single experiment by matching data sets with protein databases. Since in its essence this is a probabilistic process, there is a chance that the answer may be wrong. This can have serious implications, since the identified proteins often form the starting point for a series of new experiments.

In his article Dr. Jeroen Krijgsveld points out that in order to improve the confidence in protein identification, quality control routines should be implemented in the analysis of complex mixtures of proteins. "Methods for improving the quality of mass spectrometry outcomes are available, but have not gained much attention so far," says Krijgsveld. "It is an upcoming notion, essential for reliable interpretation of data sets."

Krijgsveld advocates the use of reversed databases. This helps to determine the rate of false positive identifications in large or even moderately sized experiments. He demonstrates that further improvement can be reached by increasing the mass accuracy. While a false positive rate of 5 to 10% is usually no exception, using an LTQ-FT mass spectrometer, a value of 0.2% can be reached for complex samples such as embryonic stem cells.

Apart from the robust identification of proteins, their quantitation by the use of stable isotopes provides an additional level of information that can be used to study the dynamics of biological systems. Using isotopically enriched media, detailed information can be delivered on the proteins present and their quantities in a sample containing hundreds of proteins. Full labelling with the stable isotope 15N has been reached even with model organisms such as the fruit fly *Drosophila*.

However, accurate analysis of complex proteins samples is not a technology that stands on its own. "Interaction with biologists, statisticians and bioinformaticians is vital for careful design of the experiments and for subsequent biological validation of the results."

Research Theme NPC1: Accurate comparative proteomics and protein quantitation

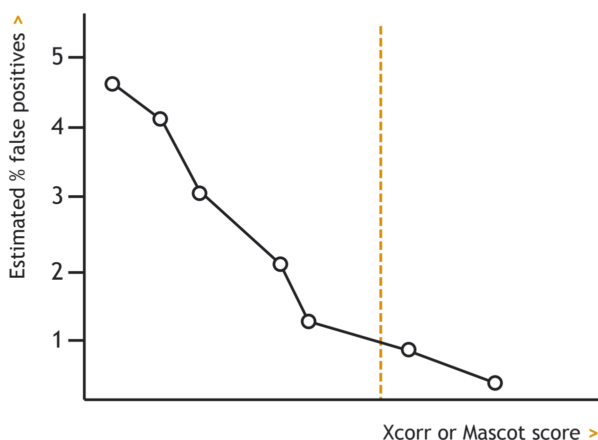


Figure 1 | **Limiting the false positives.** Simulation of the relation between probability scores used for protein identification and false positive rate. The dashed red line indicates the cut-off score that should be used to limit the false positive rate to 1%.

match. From a typical LC-MS analysis experiment, thousands of peptide fragmentation patterns are often matched to databases, each containing thousands of protein sequences. Several search algorithms have been developed to do this in an automated fashion, of which Sequest and Mascot may be the best known [2]. Generally, the outcome is coupled to a probability score reflecting the quality of the match, and a cut-off score is then used to discriminate between solid and weak identifications. Thus, in its essence protein identification by mass spectrometry is a probabilistic process with the inherent chance that the answer may be wrong.

Although this is a straightforward procedure, the crucial point is the robustness of the list of identifications that is handed to the biologist. After all, it is the biologist who will use the data as a foundation for a series of time-consuming experiments involving *in vivo* studies, producing antibodies, performing immunoprecipitations etc. The true danger lies in the inclusion of false positive identifications, and it is evident that they must be reduced to the minimum. In this respect it is surprising that, with large-scale approaches emerging abundantly, one hardly ever encounters any form of quality control of the data produced, and that this is not demanded by journals publishing them [3].

Quality control One possibility for implementing quality control is to go through the data manually to check peak annotation of individual spectra. Although this always is a good idea, it cannot be reasonably expected considering the sheer size of the datasets. The use of reversed databases can be of great help in determining the rate of false positive identifications in large or even moderately sized experiments.

The idea of using randomised or reversed databases was introduced by Moore et al [4] and further extended by Gygi et al [5]. They assumed that any protein identification produced from a search in a database consisting of nonsense DNA-sequences reflects a false positive identification in a real database. Since this assumption is only valid if such a decoy database is of the same size as the true database and contains proteins of the same average composition, the idea of reverse databases was used where every sequence had been swapped in orientation. The false positive rate could then be calculated from the number of identifications in reversed and true databases. More conveniently, one can determine the cut-off score (Mascot score, or Xcorr in Sequest) in this way such that this rate drops to an acceptable level (see Figure 1). This will generally be at the cost of the number of identifications, but will increase the confidence in the results considerably.

False positives The next question is how one can minimise the false positive rate while keeping the number of identifications at the desired level. It is therefore important to be aware of three factors that have an effect on the false positive rate: the size of the database that is searched, the size of the experimental dataset, and the mass accuracy of the experimental data. Generally, only limited influence on the first two

aspects is possible: databases often contain the entire proteome of an organism, and average datasets contain hundreds to thousands of spectra. It is therefore mass accuracy, and inherently the choice of a certain mass spectrometer, that can have profound impact on the results. The lab at the department of Biomolecular Mass Spectrometry in Utrecht routinely analyses digests of gel-separated protein samples of considerable complexity such as extracts of cells or organelles. Using an ion-trap a mass accuracy of approximately 1 Da can be achieved for the precursor ion. In such analyses it is no exception to find false positive rates between 5 and 10% when using standard cut-off scores. Slightly depending on the experiment, it is questionable whether this is generally acceptable. The recent acquisition of the LTQ-Fourier-Transform (FT) mass spectrometer has made clear the true added value of mass accuracy for protein identification.

One of the first large scale analyses running on this instrument is an NPC-project studying embryonic stem cells. The primary interest is to identify proteins in mouse and human stem cells, both before and after initiation of differentiation, in an effort to identify those proteins that are specific for either stage. Using the LTQ-FT in conjunction with a nano-flow LC system, a mass accuracy of 2 ppm is routinely achieved. By virtue of this mass accuracy and by using rigorous filtering conditions we were able to identify approximately 1700 unique proteins in each of the four cell types with a false positive rate of no more than 0.2%. It is noteworthy that this includes proteins identified by a single peptide that are usually disregarded, but can now be included as confident identifications. This dataset now provides a rich and dependable source for further biological experiments. Our main interest now is in the 180 proteins that were identified in both human and mouse embryonic stem cells, but not in differentiated cells of either species.

The extra dimension In many experiments, protein identification alone can provide important clues to biological processes. When those proteins can also be quantified, the impact of such data is further increased. This is especially valuable in comparative studies, for instance between diseased and healthy people, but also in sampling time series.

Using stable isotope labels (e.g. 15N or 13C) in proteins, mass spectrometry has the power to separate and identify differentially labelled species. This means that proteins in one sample tagged with a heavy isotope can be qualitatively and quantitatively distinguished from a control sample without such a label. In the mass spectrum, labelled and unlabelled peptides appear as peak pairs that can be quantified by their relative intensities (see Figure 2).

In order to introduce mass labels into proteins several chemical tagging strategies have been suggested [6]. Metabolic labelling is one of the preferred strategies, which implies the culturing of cells in isotopically enriched media. Clear advantages over other strategies are that labelling goes to

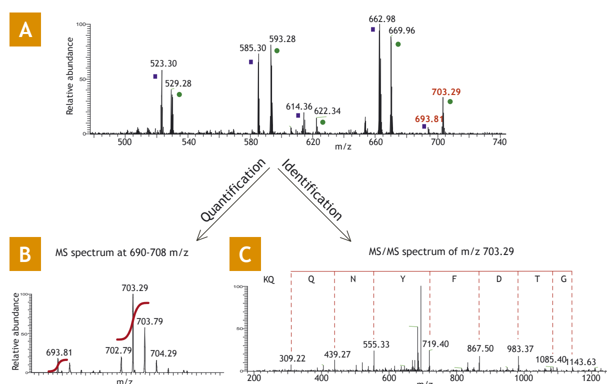


Figure 2 | **Protein quantitation by stable isotope labelling.** A mixture of labelled and unlabelled peptides is analysed with mass spectrometry: (A) The peptides are detected as peak pairs in the mass spectrum. (B) The peptide can be identified from the fragmentation pattern of either of these peaks. (C) Integration of the area under each of the peaks will deliver their relative abundance.

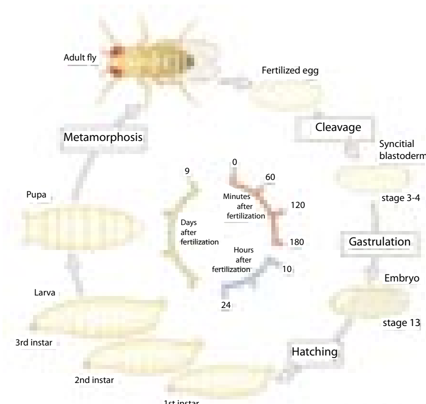


Figure 3 | **The life cycle of the fruit fly Drosophila melanogaster.** *Drosophila* development is a tightly orchestrated process, with crucial events executed at well-defined points. For instance, during the first hours after egg deposition, the embryo switches from a maternally deposited protein expression program to an embryonic program preparing the embryo for further developmental stages. In the experiment described, unlabelled embryos and embryos labelled with 15N were collected at 90 and 45 min, respectively, for a quantitative mass spectrometric analysis.

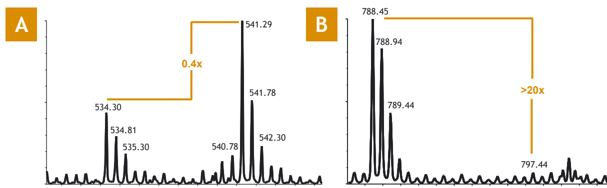


Figure 4 | Proteins changing in expression during fruit fly development. The expression of proteins changes during development. Peaks are shown of labelled (heavy) and unlabelled peptides (light), isolated from embryos at 45 and 90 min after egg deposition, respectively. Panel A shows a peptide of Me31, a protein that is down-regulated by a factor of 0.4 during this 45-min interval. Panel B shows a peptide of motor myosin, that is virtually absent at 45 min and appears strongly after 90 min.

completion for all proteins, and that labelled and unlabelled cells can be mixed in the earliest possible stage of sample handling. Thereby possible introduction of errors during mass spectrometry analysis experiments is avoided. Full labelling with 15N has been achieved for various unicellular organisms, but more importantly also for the worm *C. elegans* and the fruit fly *Drosophila* [7].

Fruit flies under development Currently the above described techniques are being applied in a large-scale project focussing on fruit fly development, and more particularly in the early developmental events (see Figure 3). To this end labelled and unlabelled flies were collected from discrete stages at 45 and 90 min after egg deposition. The analysis was performed by LC-FTMS using 60-cm capillary columns to ensure ultimate peptide separation. Parallel to protein identification, proteins were quantified using software partially developed in house. Among the majority of the proteins that, as expected, do not change in abundance, others show clear up or down regulation.

Two examples are depicted in figure 4, where the maternal protein Me31 decreases by a factor of 0.4 over the 45-min interval. Motor myosin is a protein that is virtually absent in the early stage (labelled) and is then strongly up regulated (unlabelled). The challenge now will be to fit all proteins that are changing into a model, so that predictions can be made on the pathways that are turned on or off and that may be crucial for the developmental process. These examples show that both drastic and subtle changes can be detected.

In conclusion, advanced mass spectrometry, combining high-confidence identifications with quantitative information on protein expression levels, provides an excellent tool to get a grasp on the dynamics of biological systems. Furthermore, it allows the careful design of ensuing biological experiments to test functional implications.

References

1 Aebersold, R. and Mann, M. (2003) Mass spectrometry-based proteomics. Nature 422 (6928), 198-207
2 Sadygov, R.G. *et al.* (2004) Large-scale database searching using tandem mass spectra: looking up the answer in the back of the book. Nat Methods 1 (3), 195-202
3 Baldwin, M.A. (2004) Protein identification by mass spectrometry: issues to be considered. Mol Cell Proteomics 3 (1), 1-9
4 Moore, R.E. *et al.* (2002) Qscore: an algorithm for evaluating SEQUEST database search results. J Am Soc Mass Spectrom 13 (4), 378-386
5 Peng, J. *et al.* (2003) Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. J Proteome Res 2 (1), 43-50
6 Julka, S. and Regnier, F. (2004) Quantification in proteomics through stable isotope coding: a review. J Proteome Res 3 (3), 350-363
7 Krijgsveld, J. *et al.* (2003) Metabolic labeling of C. elegans and D. melanogaster for quantitative proteomics. Nat Biotechnol 21 (8), 927-931

Jeroen Krijgsveld, PhD
Utrecht University
Dept. Biomolecular Mass Spectrometry
Sorbonnelaan 16
3584 CA Utrecht
The Netherlands
T +31 30 253 99 74
j.krijgsveld@chem.uu.nl
www.chem.uu.nl/bioms

Summary

With all the facets of proteomics that have been constantly improved over the past couple of years such as throughput, sensitivity, accuracy, and specificity, proteomics has matured to a stage where it is no longer a technology that stands on its own. A proteomic experiment does not start with a tryptic digestion, and is not finished when the list of identified proteins has been produced. Rather, before and after these 'core activities', interaction with biologists, statisticians and bioinformaticians is vital for subsequent biological validation of the results. This paper therefore argues for the implementation of a quality control routine in the analysis of complex mixtures.

In this respect reverse databases are highly useful for estimating the false positive rate in a dataset, allowing one to take measures to lower this rate if needed, for instance by choosing a higher cut-off score. This will be highly appreciated by the collaborating biologists who might otherwise be sent astray by erroneous identifications.

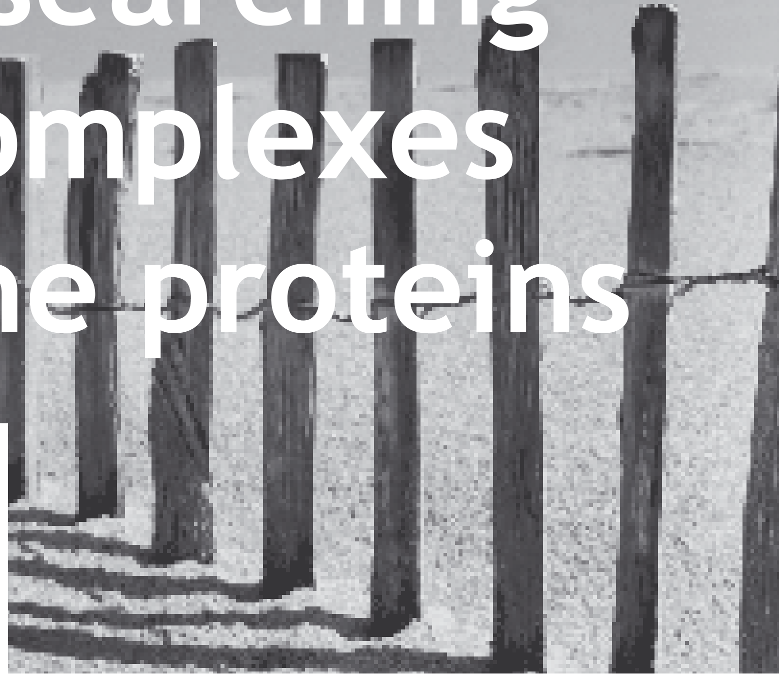
In addition, detailed information of hundreds of proteins in a sample, both qualitatively and quantitatively, can directly be delivered by stable isotope labelling. However, its application is not straightforward and necessitates careful design of the experiment by all parties involved. But, if one succeeds, integration of both types of data will put their biological impact at a significantly higher level.

summary

Ana A. Arteni, Roman Kouril and Egbert J. Boekema §

egbert boekema

systematic searching for novel complexes of membrane proteins



Membrane proteins are harder to characterise by proteomic analysis than water-soluble proteins. However, by combining mass spectrometry with single particle EM it is possible to discover novel and even larger protein complexes rather than the usual single and smaller ones. A strategy is proposed and demonstrated for systematic characterisation of novel membrane proteins.

Over the last 20 years, single particle analysis of electron microscopy (EM) projections has become a well-established technique to obtain structural information about large biomacromolecules at a resolution of 10–20 Å [1]. Single particle averaging is simple if applied to negatively stained specimens with a mass between about 300 kDa and 2000 kDa, and since it is possible to process several thousands of projections within a few days, yielding 2D projection maps of at least 20 Å resolution. The statistical analysis and classification procedures used in single particle analysis are often effective in the sorting of different projection views originating from different conformations or subunit compositions.

A proteomics analysis of transmembrane proteins is often focussed on the discovery of novel protein components or on the study of modifications by phosphorylation. Such a systematic analysis is based on one dimensional or multidimensional protein/peptide separation depending on the complexity of the sample followed by extensive (tandem) mass spectrometric analysis. In comparison with sequence tags and genomic databases, the presence of fully novel or multiple, closely related proteins can be demonstrated. Studies to test this approach have been carried out on several types of membranes, such as mitochondrial membranes from yeast. This paper focuses on photosynthetic membrane proteins from oxygen evolving organisms.

Photosynthetic complexes During the analysis of a large set of 16,600 EM projections of photosynthetic PS2 complexes, purified from chloroplast grana membranes from spinach, a strong heterogeneity of particles was noticed due to a variable binding of peripheral subunits around the PS2 core complex [2]. As a result, ten protein complexes with a different subunit composition or arrangement were resolved. However, in this analysis, which consisted of several cycles of alignments, multivariate statistical analysis and classification of about 25% of the projections could not be assigned to any of these PS2 particles [2]. One triangular shaped protein contaminant with a diameter of about 200 Å was present in about 200 copies. By averaging these projections, the contaminant could be designated as a multimer of seven copies of trimeric light-harvesting complex LHCII [3].

This assignment was possible for two reasons. First, a high-resolution structure of LHCII trimers was solved by electron microscopy. Even at the 20 Å level the strong similarity be-

What this research is about:

Discovering novel membrane proteins

Electron microscopy (EM) is a useful technique for studying the structure of proteins with a mass above about 200 kDa. It produces 2D-projection views of such proteins. By averaging these projections with an image analysis technique, such as single particle analysis, structural information of large biomacromolecules is obtained at a resolution of 10-20 Å or even better.

"Although single molecules usually have a rather random orientation in a microscope sample, with the help of statistical analysis and classification programs different types of views can be separated and averaged," says Egbert Boekema. "This technique enables one to detect projections from mixtures of biomolecules, even if a contaminant is present with a frequency of about 0.2%. This opens the way to look for all the larger components of a certain type of membrane."

In their article Boekema and his colleagues propose a method that combines single particle EM and protein mass spectrometry. Its strategy is specially meant for membrane proteins, which are harder to characterise by a proteomics analysis than water-soluble proteins.

They demonstrate the potential of the method on the basis of Photosystem II (PS2), a multisubunit membrane protein complex, which catalyses one of the key reactions on our planet: the light-driven oxidation of water. To their surprise they were able to identify a new subunit.

Boekema: "This kind of research is generally focussed on one purified protein. But in the case that you don't know where you are looking for, you can purify endlessly. The technology is now so accurate that we can scout a sample very specifically and discover what's there. This sample could be membrane fragments, mixtures or isolated proteins. With EM we are able to detect every particle. This has never been done before, simply because we weren't able to. Since we have developed this strategy we thought: Why not do it in a systematic way?"

Research Theme NPC2: Membrane proteomics

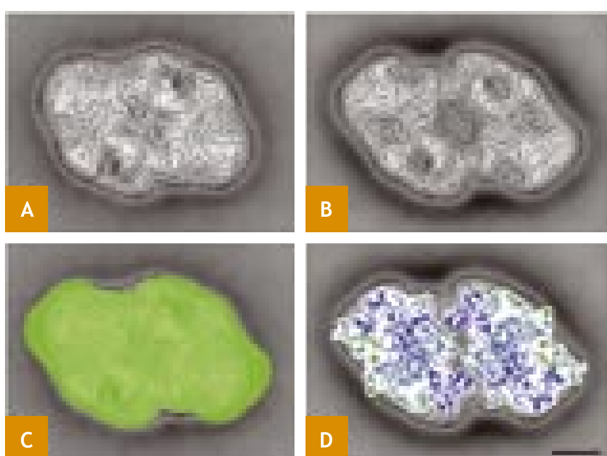


Figure 1 | **Single particle analysis of top view projections of PS2 complexes.** (A) averaged top view projection of PS2 complexes purified with a standard HPLC procedure. (B) averaged top view projection of His-tagged PS2 complexes after modified HPLC. (C) map of image of B (in green) superimposed on image A. (D) Fitting of the X-ray structure [7] on image B suggests that the additional mass belongs to two membrane-bound α-helices (in blue-green). Space bar is 50 Å

tween the contaminant and the LHCII structure was apparent. Second, it is known that single LHCII trimers which do not bind to PS2 are present in variable, but appreciable numbers within the thylakoid membrane, from which PS2 complexes were purified. However, the discovery that these trimers would be present as supercomplexes with seven LHCII trimers was a surprise.

These and other results on rather heterogeneous sets of EM projections indicate that single particle EM analysis has the potential to detect specific unknown large complexes in large data sets, and to obtain reasonably well defined 2D maps, even if such contaminants are present in small numbers. These studies also indicated that it would be likely that by checking 'rather pure' membrane protein complexes for contaminants, more of such novel structures could be found if averaging of several dozens to hundreds of projections is possible.

So if it is possible to assign such novel structures to polypeptides in a rather systematic way, the technique of single particle EM analysis can get a novel and useful application in the expanding field of proteomics.

This paper proposes a method, which combines single particle EM and protein mass spectrometry. This strategy is specially meant for membrane proteins, which are harder to characterise by a proteomics analysis than water-soluble proteins and aims to discover novel and larger complexes rather than the single smaller proteins [4].

PS2 analysis To test the possibilities of combining single particle EM with protein mass spectrometry two batches of PS2 from the cyanobacterium *Thermosynechococcus elongatus* are compared. One was purified by application of three successive chromatography steps, the other by just one step, by taking advantage of an introduced His-tag on one of the subunits. This was done in collaboration with M. Rögner's group in Bochum Germany.

PS2 is a multisubunit membrane protein complex, which catalyses one of the key reactions on our planet – the light-driven oxidation of water. It is a dimer with over 25 subunits: PsbA to PsbZ and many others. About half of them are small intrinsic membrane proteins with a molecular weight below 10 kDa.

In contrast to the standard PS2, which was purified by a standard HPLC procedure, gel electrophoresis shows that the His-tagged PS2 complex has one additional mass of about eight kDa. Two sets of about 15,000 projections of dimeric PS2 complexes were analysed. By comparing the two dimers, it appears that they are highly similar, except for one site at the outer tips outlined in the difference image (see Figure 1C). This tip is larger in the particles with a His-tag. Comparison of these EM data with a 3.7 Å X-ray structure from PS2 indicates that two membrane-spanning α-helices at the tip of the

projection are responsible for this difference (see Figure 1D). The PS2 complexes were then directly applied to MALDI-TOF analysis in order to determine the entire mass of the small subunits. With this method we could assign a single mass in the range of 7000 to 8000 m/z to subunit PsbZ in the sample of the His-tagged PS2 complex; this mass is not present in the standard PS2 preparation. This experiment shows that mass differences between membrane proteins on the order of 3% are still detectible, as shown for the PsbZ subunit of PS2.

L-shaped contaminant In the PS2 preparations studied above one contaminant appeared. L-shaped particles were present with an abundance of 2% and 10% in the batches of PS2 with the His-tag and the ones without, respectively. Image analysis indicated that this L-shaped particle appeared in two dominant views (see Figure 2A). Because projections in both views have the same overall dimensions, they likely come from the same particle and thus only have a different handedness in respect to the carbon support film.

Because Complex I is, to our knowledge, the only known major membrane complex which is L-shaped, we compared projections with those from His-tag purified Complex I (see Figure 2B), which confirms our hypothesis. The cyanobacterial Complex I, more commonly known as NDH-1, has a much smaller vertical arm than all systems studied, such as the one from *Arabidopsis* (see Figure 2C) [5]. The reason is the lack of three genes responsible for a mass of 150 kDa [6]. This mass holds the NADH dehydrogenase activity and is sometimes disrupted in *Arabidopsis* (see Figure 2D). Surprisingly, in a set of about 25,000 projections of a *Synechosystis* cyanobacterium, about 0.2% of the particles have an additional mass such as the one of *Arabidopsis* (in collaboration with E.M. Aro's group in Turku, Finland). This indicates that there must be structural homologous proteins. The next step is to figure out which proteins are responsible for this mass in a combined EM and mass spectrometry approach.

Scheme for systematic analysis The following strategy specially focuses on a characterisation of membrane proteins by proteomics. The idea is to study all the larger proteins. The easiest way to obtain them would be a simple one-step size-exclusion chromatography purification to remove the abundant smaller proteins and fragments. The first step in the strategy is to detect novel 2D structures present in a specific protein fraction of a detergent-solubilised membrane by single particle EM. This could be a specialised type of membrane, such as the thylakoid membrane from photosynthetic organisms. The second step would be to obtain a batch from this membrane in which single particle EM indicates that the complex of interest is more abundant or largely absent. This could be achieved by applying further different partial purification procedures, like sucrose gradient centrifugation or dialysis with membranes with a high cut-off.

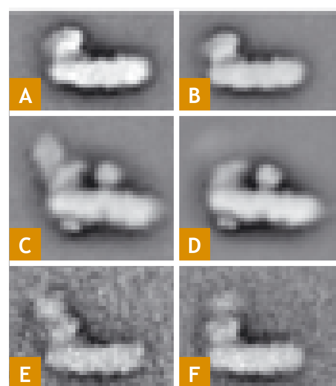


Figure 2 | L-shaped particles (A) Analysis of L-shaped particles from *Thermosynechococcus elongatus*, tentatively assigned to Complex I. (A) sum of 1962 projections. (B) Similar sum from His-tag purified NDH-1 from *Cynechosystis*. (C-D) Complex I and its major fragment from *Arabidopsis* (E-F) unexpected large NDH-1 particles from *Synechosystis*.

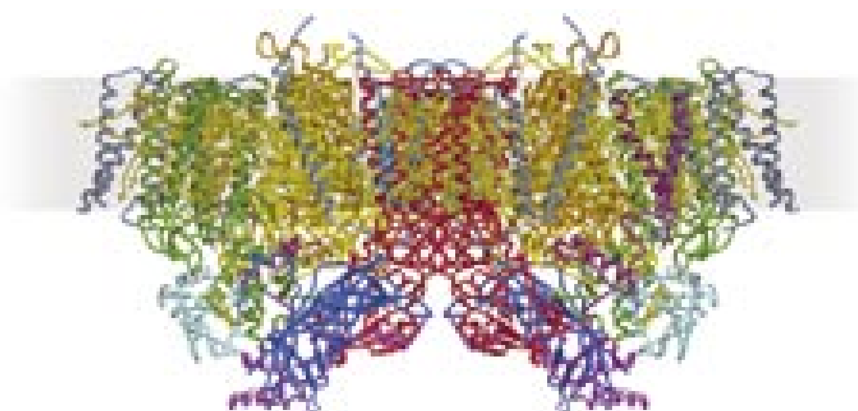


Figure 3 | Photosystem II, a multisubunit membrane protein complex. X-ray diffraction model from Photosystem II (PS2), seen in side view position and determined by Prof. So Iwata (Imperial College, London, published in 2004, Science, Vol 303 P 1831-1838). The position of the membrane is indicated in grey. See also http:/www.esi-topics.com/ for a discussion for a wider audience.

Third, if differences in the frequencies of observed novel structures complexes are detected by single particle EM, SDS-PAGE patterns from the various purified membrane protein batches should be inspected for major differences. Thereafter, mass spectrometry (MALDI-TOF and other techniques) should be used to assign protein bands with a variable intensity. Since the genomes of many organisms have already been sequenced, an unambiguous assignment could be achieved in most cases. If necessary additional *de novo* synthesis could be of help. Further studies could also be undertaken by subsequent purification of the novel structures to homogeneity.

The major advantage of EM over many other biophysical methods is certainly the detection of single molecules. Single particle EM should take full advantage of this. If so, application of the proposed scheme can yield interesting complexes or transient supercomplexes in many membranes to be discovered.

References

1. Frank, J. (2002) Single-particle imaging of macromolecules by cryo-electron microscopy. *Annu. Rev. Biophys. Biomol. Struct.* 31, 309-319.
2. Boekema E.J. et al. (1999) Multiple types of association of photosystem II and its light-harvesting antenna in partially solubilized photosystem II membranes. *Biochemistry* 38, 2233-2239.
3. Dekker, J.P. et al. (1999) Heptameric association of light-harvesting complex II trimers in partially solubilized photosystem II membranes. *FEBS Lett.* 449, 211-214.
4. Arteni, A.A. et al. (2005) Single particle electron microscopy in combination with mass spectrometry to investigate novel complexes of membrane proteins. *J. Struct. Biol.* 149, 325-331.
5. Dudkina N.V. et al. (2005) Structure of a mitochondrial supercomplex formed by respiratory chain complexes I and III. *Proc. Natl. Acad. Sci* 102, 3225-3229.
6. Friedrich, T. et al. (1995) The proton-pumping respiratory complex I of bacteria and mitochondria and its homologue in chloroplasts. *FEBS Lett.* 367, 107-111.
7. Ferreira, K.N. et al. Science (2004) Architecture of the photosynthetic oxygen-evolving center. *Science* 303, 1831-1838.

Research group Authors and some other members of the electron microscopy group: left to right: Natalia Dudkina, Ana Arteni, Gert Oostergetel, Roman Kouril, Egbert Boekema and Mihaela Folea.

§

Egbert Boekema, PhD
University of Groningen
Department of Biophysical Chemistry, GBB
Nijenborgh 4
9747 AG Groningen
The Netherlands
T +31 50 363 42 25
boekema@chem.rug.nl

Summary

Single particle analysis of electron microscopy (EM) projections has become a well-established technique to obtain structural information about large biomacromolecules at a resolution of 10–20 Å. It is also possible to detect projections from mixtures of biomolecules, even if a contaminant is present with a frequency of about 0.2%. This paper proposes a strategy to combine single particle EM and protein mass spectrometry to look for all the larger components of a certain type of membrane.

To demonstrate the possibilities the photosynthetic membrane proteins from oxygen evolving organisms are taken as an example. During the analysis of a large set of 16 thousand EM projections of photosynthetic PS2 complexes, purified from spinach, a strong heterogeneity of particles was found around the PS2 core complex. About 25% of the projections could not be assigned to any of the PS2 particles. One of the contaminants appeared to be a triangular shaped protein of 200 Å in about 200 copies. By averaging these projections, the contaminant could be designated as a multimer of 7 copies of trimeric light-harvesting complex LHCII. A surprising discovery.

This indicates that single particle EM analysis has the potential to detect specific unknown large complexes in large data sets, even if such contaminants are present in small numbers. These studies also imply that by checking 'rather pure' membrane protein complexes for contaminants more novel structures could be found. Since it is seems possible to assign such novel structures to polypeptides in a rather systematic way, the technique of single particle EM analysis deserves its place as a novel and useful application in the expanding field of proteomics.

summary

roland pieters

old concepts and new methods detect galectins complex biological sampl

Interfering with protein-carbohydrate interactions has great potential in medicine. Especially galectins are of interest because of their role as a biomarker for cancer. The design of a probe to detect galectins is a crucial step towards the diagnosis and prognosis of malignant tumors. The notoriously weak interaction between the galectin and the probe is a problem. Several timely concepts and some revisited old concepts were brought together to overcome this hurdle.

Carbohydrate recognition is of great importance in biological processes. Most cells are covered by carbohydrates that are well-positioned for their involvement in cell-cell recognition phenomena as e.g. in the immune system, leukocyte recruitment and inflammation, cancer and metastasis, bacterial adhesion and infection, blood coagulation and cell differentiation. The proteins that are involved in these biological processes are therefore important players. These could be proteins that bear the carbohydrate structures, for example the glycosylated proteins, or they could be proteins that recognise these structures, such as lectins. Among the lectins the galectins (see Figure 1) are a group with particular relevance as biological marker for, amongst others, cancer. In many laboratories worldwide research on the biological roles of the fourteen galectins is the subject of intense study. Their involvement in the immune system and in cancer, however, is clear.

The goal of this work is to create an easy method that allows the detection of the levels of the various galectins in a complex biological sample with great sensitivity and accuracy. Such a method would have applications e.g. in the diagnosis and prognosis of malignant tissue.

Molecular probes are being designed, synthesised and tested to achieve this goal. These probes are multifunctional molecules that are capable of specifically binding the galectins and covalently attaching themselves to them. Once labelled, the other end of the molecule features functionality that allows the visualisation of the labelled proteins. Our group has developed expertise in the interference with protein-carbohydrate interactions in general and the interference with galectins in particular. The ability to synthesise a variety of carbohydrate-containing candidate probes and test their efficacy in gels is crucial for this work. Mass spectrometry is also an indispensable tool to verify the covalent attachment of the probe to the desired proteins.

Capture The galectins all share a binding specificity for the carbohydrate galactose at the end of carbohydrate chains and are linked in a so-called ß-fashion to the adjacent sugar. To take advantage of this recognition we set out to design a probe that contained this motif.

Capturing the protein by creating a covalent bond between probe and bound protein was the biggest challenge. The

What this research is about:

Click chemistry appears to be a useful tool for galectin quantification in plants

Galectins are a special class of protein that can recognise carbohydrates in cells. As protein-carbohydrate interactions play a role in cell-cell recognitions, galectins are particularly important as biological markers for cancer. Detection of galectins in tissue therefore has potential in diagnosis and prognosis of cancer. Dr. Roland Pieters has designed a probe molecule that binds to galectins and facilitates detection of this protein in tissue.

For detection, a probe molecule should be covalently attached to the target. Galectins do not bind covalently to carbohydrates, therefore a trick was needed for the probe to attach to the galectin. Pieters explains: "We designed a probe molecule with an attached photo-affinity label which does attach itself covalently to the galectin. After the protein was captured the opposite end of the probe could be functionalised with a fluorescent label to visualise the galectins in a mixture with other proteins."

Pieters and his group developed the required chemistry to synthesise the probes using new techniques and modifying old methods. The new use of microwave irradiation, which highly accelerated important reactions, was crucial for the selective addition of the photo-affinity label. The attachment of the fluorescent label was possible because of the application of a recently discovered copper-catalysed cycloaddition reaction called 'click chemistry'. The work underscores the usefulness of organic synthesis in the field of proteomics since the type of probes described in this article will facilitate the study of important subsections of the proteome.

Research Theme NPC3: Protein post-translational modification

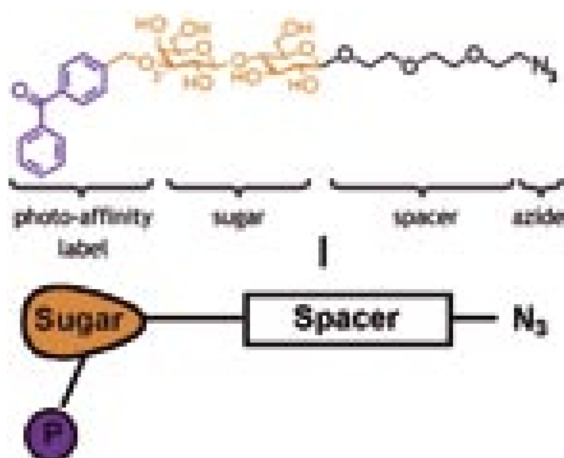


Figure 1 | Crystal structure of galectin-1

galectins only bind to their target sugar by non-covalent interactions and do not interact via a covalent mechanism like enzymes do. In order to capture enzymes, advantage can sometimes be taken of the enzymatic mechanism [1]. In these cases the enzyme can perform the first addition step, but, due to the specific features of the modified substrate used, is subsequently incapable of moving forward to complete the catalytic cycle. The amount of captured protein is a measure of the amount of active enzyme, which is more relevant than determining the total amount of enzyme present which could be (partially) inactivated by natural inhibitors.

This strategy, however, could not be used for the galectins because they are not enzymes and merely bind their target sugar sequence. To overcome this we attached photo-affinity labels to the ß-galactoside. The photo-affinity label can covalently attach itself to protein residues if it comes within a short distance of protein residues when the ß-galactoside is bound in the binding groove and when it is irradiated with light of the proper wavelength. Our first attempt indicated that we placed the photo-affinity label too far away from these residues since the probe failed to capture the galectin, even though binding of the probe was shown to occur. Based on this information we studied the available crystal structures of the galectins to see where photo-affinity label placement would be more productive. The 3'-position of a lactose ligand seemed promising since the area available around this carbohydrate oxygen was

conserved amongst the galectin family.

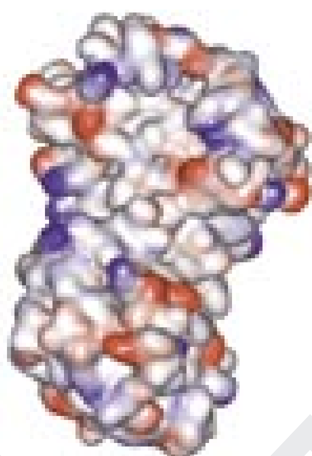
The next challenge was the synthesis of such a molecule. The carbohydrate lactose has many hydroxyl groups and we only wanted to couple the photo-affinity label to one of them. Synthetic methodology involving tin-acetals was available, but reactions were lengthy and often low yielding. Experimenting with microwave irradiation of the reaction mixture proved useful and the reaction was both faster and yielded more of the desired probe (see Figure 2). The method proved to be general for the addition of all kinds of groups to the 3-OH of a galactose unit [2].

Click chemistry The effect of the new placement of the photo-affinity label was evaluated by mass spectrometry of a galectin that was incubated with the probe and irradiated with light. The mass spectrum showed that around 70-80% of the galectin had a probe molecule covalently attached to it. With this promising result we moved ahead and tried to develop the protocol that would selectively visualise the captured galectins in a protein gel (see Figure 3). In order to do this we reacted the other end of the probe molecule, which contained an azido group, with an alkyne-linked fluorescent dye, in a highly selective coupling reaction. This coupling reaction between an alkyne and azide was recently rediscovered [3] and greatly improved by copper catalysis and yields a triazole moiety. Most importantly, the transformation can be performed in the presence of all the functional groups present in biomolecules. This so-called 'click chemistry' reaction was ideally suited for our purposes, since now the fluorescent dye did not have to be introduced into the probe until after it had already captured the protein. This greatly expands the areas of application of the probe since the bulky dyes often lead to interference with binding, have non-specific interactions or lead to reduced transport properties hampering *in vivo* work.

After some optimisation of the conditions for the 'click chemistry' a fluorescent image on a protein gel could be observed at the proper position when galectin-1 was used. When denatured galectin-1 was used, or competing galactosides were added, no such image was observed, indicating that binding of the probe to the protein was required.

Finally, a sample of four proteins was tested that contained galectin-1 and -3 and two non-galectins. After using the protocol involving probe incubation, irradiation and 'click chemistry' the fluorescent image on the gels showed only the two galectins, while the Coomassie stained gel showed all four proteins, indicating that indeed only the galectins were labelled.

Scratching the surface The identification of the biological and medical importance of the galectins is still only in its infancy. Despite the hundreds of publications on galectins their roles are so varied, even between intracellular processes and extracellular processes, that we are just scratching the surface. It is becoming increasingly clear that galectins play an important role in cancer and particularly in the metastasis of cancer. Improved and simpler methods to get quantitative insights into the presence of galectins in tissues are of growing importance. The work that is described above and that we also recently published [4] showed that it is possible to selectively label and visualise galectins in the presence of other proteins even though protein-carbohydrate interactions



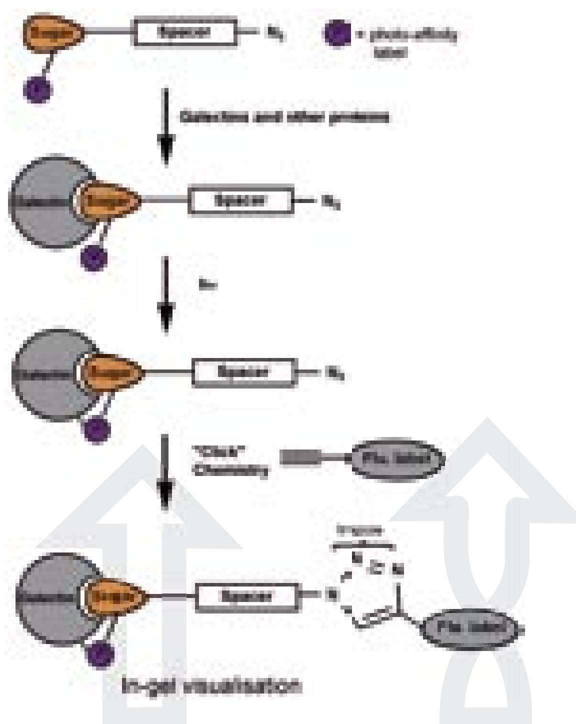


Figure 2 | **Probe.** Chemical structure and schematic representation of the galectin-selective probe.

Figure 3 | **Incubation, irradiation and click chemistry**. Protocol for selective visualisation of galectins in a mixture of proteins using the probe as represented in Figure 2.

are notoriously weak and also despite the fact that galectins are not enzymes, thus no chemical mechanism is available to take advantage of.

The present account shows that several timely concepts such as microwave-assisted chemistry and click chemistry, and also revisited old concepts such as photo-affinity labelling had to come together in order achieve the desired results. Examining crystal structures was also an important factor.

Furthermore, we have shown that in the specific case of the galectins progress can be made in the design and synthesis of probe molecules that selectively capture members of this segment of the proteome. The work generates optimism that optimised reagents of this type may indeed find application for use in diagnosis and prognosis. In order to improve the first generation that is described here, we are currently taking advantage of multivalency, a common method in nature that enhances binding affinities. By linking multiple copies of the sugar ligands together, we have previously shown that enhanced binding to the galectins results [5]. The probes based on this are in development and have already shown promise with respect to sensitivity and selectivity.

References

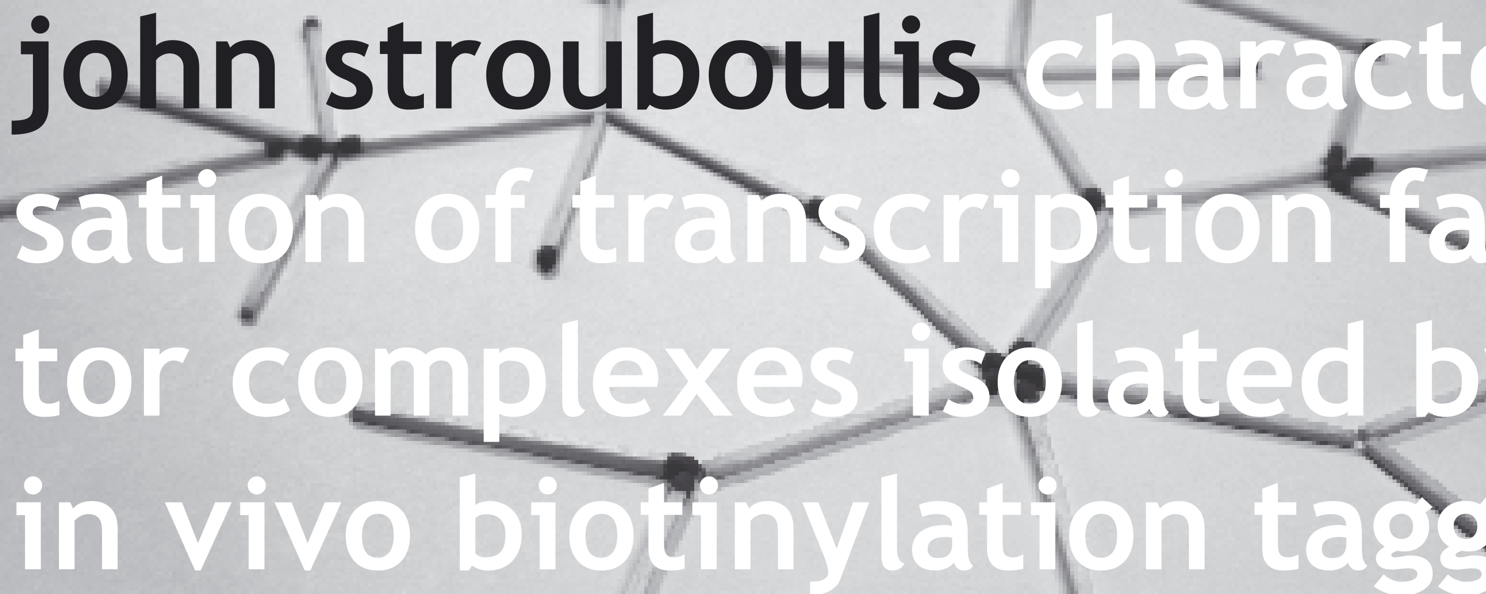
1 Jessani, N. and Cravatt, B. J. (2004) *Curr. Opin. Chem. Biol.* 8, 1-6.
2 Ballell, L. et al. (2004) *Tetrahedron Lett.*, 45, 6685-6687.
3 a) Rostovtsev, V. V. et al. (2002) *Angew. Chem., Int. Ed.*, 41, 2596-2599; b) Tornoe, C. W. et al. (2002) *J. Org. Chem.* 2002, 67, 3057-3069.
4 Ballell, L. et al. (2005) *ChemBioChem*, 6, 291-295.
5 a) Vrasidas, I. et al. (2003) *Org. & Biomol. Chem.* 1, 803-810; b) André, S. et al. (2001) *ChemBioChem.* 2, 822-830.

Roland Pieters, PhD
Department of Medicinal Chemistry
Utrecht Institute for Pharmaceutical Sciences
Utrecht University
P.O. Box 80082,
3508 TB Utrecht
T +31 30 253 69 44
r.j.pieters@pharm.uu.nl

Summary

Galectins are an important class of carbohydrate binding proteins that play roles in the immune system and most prominently in cancer. In our laboratories we are developing synthetic probe molecules as tools that enable the quantification of the galectins in complex protein mixtures such as tissue samples. These tools should be useful for the determination of the diagnosis and prognosis. Successful probes were prepared that contain the carbohydrate ligand with an attached photo-affinity label. The latter was able to covalently attach itself to the protein. After the protein was captured the opposite end of the probe could be functionalised with a fluorescent label that allowed the selective visualisation of galectins in a mixture with other proteins. The chemistry required to synthesise the probes involved new use of microwave irradiation which highly accelerated important reactions. Furthermore, the attachment of the fluorescent label was possible because of the application of a recently discovered copper-catalysed cycloaddition reaction, now usually referred to as 'click chemistry'. The work underscores the usefulness of organic synthesis in the field of proteomics since the type of probe described in this article will facilitate the study of important subsections of the proteome.

summary



john strouboulis characterisation of transcription factor complexes isolated by in vivo biotinylation tagg

GATA-1 is an essential hematopoietic transcription factor for the differentiation of erythroid cells. GATA-1 functions include upregulation of erythroid genes, suppression of cell proliferation and cell cycle arrest and repression of transcription programs associated with earlier stages of hematopoiesis or alternative lineages. To investigate how both functions, activation and suppression, are executed in erythroid cells by GATA-1 and other transcription factors, a simple method has been applied for the single-step isolation of transcription factor complexes. The isolated complexes allowed us to study and characterise the interaction of GATA-1 with other factors.

Hematopoiesis is the process responsible for the generation of the different blood cells and has often served as a model for understanding the basis of cellular commitment and differentiation. Hematopoiesis proceeds from multipotential, self-renewing hematopoietic stem cells (HSCs) to a number of progressively divergent and increasingly defined lineages, each lineage giving rise to one type of mature blood cell. The differentiation of distinct hematopoietic lineages is the result of specific transcription programs regulated by a number of essential transcription factors.The key functions that transcription factors play in hematopoiesis have been revealed by their targeted gene inactivation, which often results in the arrest of differentiation in one, or more, lineages [1]. Recent evidence has indicated that the 'plasticity' associated with HSCs is reflected at the molecular level by the low level expression of genes associated with differentiated hematopoietic lineages[2]. This led to a model of HSCs being transcriptionally "primed" for differentiation along different lineages. Thus, transcription factor function in lineage differentiation involves not only the upregulation of the lineage specific genes, but also the repression of the transcriptional 'promiscuity'[2]. However, little is known about how transcription factors execute both (i.e. activation and repression) of these functions, in parallel, in the same cells. In this study a simple but efficient approach based on the *in vivo* biotin tagging of specific transcription factors for the direct isolation and characterisation of protein complexes has been adapted, with the aim of identifying how the protein partners' transcription factors interact within mediating their diverse functions in differentiation.

Lineage differentiation The work focussed initially on the characterisation of GATA-1 protein complexes from red (or erythroid) blood cells. GATA-1 is a DNA sequence-specific zinc finger transcription factor that is essential for the differentiation of erythroid, megakaryocytic, eosinophil and mast cell lineages [3]. Significantly, enforced expression of GATA-1 in other hematopoietic lineages reprogrammed them towards the erythroid and megakaryocytic lineages, with the concomitant repression and upregulation of respective lineage-specific marker genes. In erythropoiesis, GATA-1 functions have been

What this research is about:

Biotin provides a strong handle for pulling out transcription factor GATA-1

The different blood cells in our bodies are generated from stem cells in the bone marrow by a process called hematopoiesis. Hematopoiesis is mediated by transcription factors, which regulate specific 'batteries' of genes by turning them on and off in response to specific cues. The particular combination of genes regulated by a specific transcription factor complement determines which differentiation pathway is taken and hence the type of blood cell generated. However, this remains a poorly understood process. GATA-1 is a transcription factor that is essential for the differentiation of erythroid (red blood) cells. Without GATA-1 no erythroid cells are formed.

GATA-1 functions, like those of many other transcription factors, involve the activation and suppression of different genes. For example, GATA-1 activates genes required for red blood cell functions and represses genes associated with other blood cell types. "Our main question is how GATA-1 can execute these very different functions in the same erythroid cell," John Strouboulis says. Previous research indicated that GATA-1 forms complexes with other transcription factors, which together help in executing these functions. Strouboulis: "To study these complexes we applied a method for their easy purification. Essentially, a small piece of protein was fused to mouse GATA-1 and introduced in mouse red cells. This extra piece of protein is recognised by the bacterial enzyme BirA, also introduced in the red cells, which adds to it a small molecule called biotin. The addition of biotin provides a very strong handle for pulling out GATA-1, and its interacting complexes, from red cells and identifying them by sophisticated methods, without the need for laborious and costly pre-purification steps."

Further analysis revealed that GATA forms at least five distinct complexes of which three had not previously been identified. The complexes perform distinct functions (repression and activation) on different sets of target genes during red blood cell formation. The biotin tagging method should be helpful in characterisation of other transcription factor complexes. This will lead to the elucidation of the regulatory networks controlled by these factors in hematopoiesis.

Research Theme NPC4: Protein networks

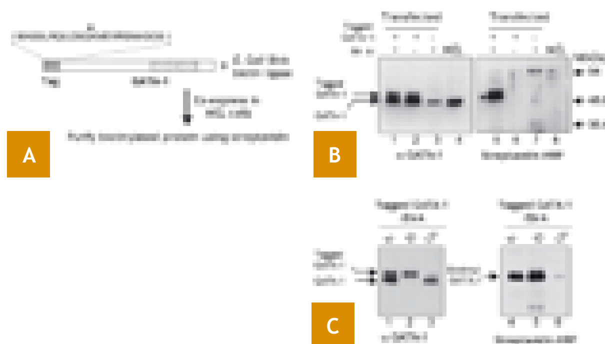


Figure 1 | (A) Scheme for the specific biotinylation of tagged GATA-1 by BirA biotin ligase in mouse erythroleukemic (MEL) cells. The sequence of the peptide tag fused to the N-terminus of GATA-1 is shown. The asterisk indicates the lysine residue that becomes specifically biotinylated by BirA. GATA-1 Zinc-fingers are indicated as speckled boxes. (B) Biotinylation of tagged GATA-1 in MEL cells. Left panel: Western blot with an anti-GATA-1 antibody to detect endogeneous and tagged GATA-1 proteins. Right panel: Western blot of the same extracts with streptavidin-HRP conjugate to detect biotinylated GATA-1. (C) Efficiency of GATA-1 biotinylation and binding to streptavidin beads. Left panel: Western blot using anti-GATA-1 antibody to detect binding of tagged GATA-1 to streptavidin beads. Input and unbound material are shown in lanes 1 and 3. Right panel: the same filter stripped and re-probed with streptavidin-HRP to detect the binding of biotinylated GATA-1 to streptavidin beads (lane 5). Lane 6 shows that very little tagged GATA-1 remains unbound by streptavidin. (In: Input (nuclear extract); El: Eluted material; Un: Unbound material).

associated with the upregulation of erythroid genes, the suppression of cell proliferation and, possibly, the repression of transcription programs associated with other lineages. Many GATA-1 protein interactions have been reported and include essential hematopoietic transcription factors such as TAL-1 (and its partners Ldb1 and LMO2), FOG-1, EKLF, PU.1, as well as proteins involved in chromatin remodelling and modification, such as the CBP/p300 histone acetyltransferases and the SWI/SNF complex [3]. Despite all this evidence, important questions remain as to how GATA-1, and other transcription factors, can accommodate all these functions and interactions in erythroid cells. To this end, we applied the in vivo biotinylation tagging approach to isolate and characterise GATA-1 complexes from erythroid cells.

Biotinylation tagging The biotinylation tagging approach is outlined in Figure 1A. It is based on previous work on the screening of a combinatorial synthetic peptide library for efficient biotinylation by the bacterial BirA biotin ligase, an enzyme responsible for the covalent attachment of biotin to metabolic enzymes [4]. This led to the identification of a number of short sequence tags that can be very efficiently biotinylated *in vitro* [4,5]. What makes biotinylation tagging

attractive is that biotinylated substrates can be bound very tightly by the naturally occurring proteins avidin and streptavidin (Kd = 10-15), a fact that has been widely exploited in many applications in molecular biology. In addition, there are few (mostly cytoplasmic) naturally biotinylated proteins, ensuring that non-specific background binding remains low. In this study one such short tag has been used to fuse it N-terminally to mouse GATA-1 and to co-express it with the bacterial BirA biotin ligase in mouse erythroleukemic (MEL) cells. By screening MEL cell stable transformants with streptavidin-horseradish peroxidase (HRP) conjugate we were able to observe robust biotinylation of tagged GATA-1 only in the presence of co-expressed BirA (see Figure 1B, lane 5 asterisk) [6]. Test binding of nuclear extracts to streptavidin paramagnetic beads followed by Western blotting with GATA-1 antibody and streptavidin-HRP showed that GATA-1 biotinylation under these conditions was close to 100% efficient (see Figure 1C) [6]. Larger scale binding of nuclear extracts to streptavidin beads, elution of the bound proteins followed by SDS-PAGE and staining of the gel (see Figure 2A), allowed us to compare the background binding using extracts from cells expressing BirA only (lanes 4 and 5), versus the specific binding using extracts from cells expressing biotinylated GATA-1 (lanes 2 and 3). From this it can be seen that background binding consisted of five strongly staining bands against a backdrop of fainter bands (lane 5). Importantly, the staining pattern of proteins eluted from the binding of extracts expressing biotinylated GATA-1 was significantly different to the background binding pattern showing a significant enrichment in proteins co-eluting with tagged GATA-1 (lane 3 versus lane 5). The arrow in lane 3 indicates the protein band containing biotinylated GATA-1 as

determined by mass spectrometry.

Identification of GATA-1 complexes We determined the nature of the background binding proteins by excising the entire gel lane, trypsinizing the fractionated proteins and analysing the eluted peptides by LC-MS/MS (see Figure 2B and 2C) [6]. The most abundant background binding proteins thus identified were naturally biotinylated enzymes, such as carboxylases, and associated proteins (2B). We also found background binding of abundant nuclear proteins, such as factors involved in mRNA processing, and of ribsomomal proteins (2B). By contrast, analysis of the proteins co-eluting with biotinylated GATA-1 showed a large enrichment for transcription factors and chromatin remodelling and modification proteins, suggesting specific co-purification with GATA-1 (2C) [7]. In addition, enrichment in topoisomerases was also evident and was due to their indirect co-purification with GATA-1 by virtue of their association with chromatin.By employing independent immunoprecipitation experiments using extracts from non-transfected MEL cells, we confirmed the specific interaction for some of the proteins that co-purified with biotinylated GATA-1 [7]. These included the essential hematopoietic transcription factors FOG-1, TAL-1, Ldb1 and Gfi-1b. The fact that most of these factors (but not Gfi-1b) had been previously reported to interact with GATA-1 validated our approach of biotinylation tagging for protein complex purification. In addition, we identified GATA-1 interactions with the MeCP1 and ACF/WCRF chromatin remodelling complexes for the first time. The MeCP1 complex also contains histone deacetylases (HDACs) and methyl DNA binding (Mbd) proteins and is associated with transcriptional repression. The ACT/WCRF chromatin remodelling complex has been associated with both activation and repression.

By using a two-step immunodepletion/immunoprecipitation approach we were able to further investigate the multiple interactions identified for GATA-1 by biotinylation tagging. Thus, we showed that GATA-1 forms at least five distinct complexes: with FOG-1 and, independently, with FOG-1 and the MeCP1 complex, with TAL-1 and Ldb1, with Gfi-1b and with the ACF/WCRF complex. Further characterisation of these complexes by transfection assays and chromatin immunoprecipitation (ChIP) assays suggested that the distinct GATA-1 complexes target distinct subsets of target genes [7]. Specifically, we found that the GATA-1/FOG-1/MeCP1 complex represses genes (such as GATA-2) associated with the early hematopoietic multipotential state or with alternative hematopoietic lineages (such as the eosinophilic MBP gene), the GATA-1/Gfi-1b complex represses genes associated with cell proliferation (such as myc and myb), whereas the GATA-1/TAL-1/Ldb1 and the independent GATA-1/FOG-1 complex are associated with the activation of erythroid specific genes (such as EKLF and globin).

Different tasks We have shown that *in vivo* biotinylation tagging is an efficient method for the direct purification of transcription factor complexes from nuclear extracts without

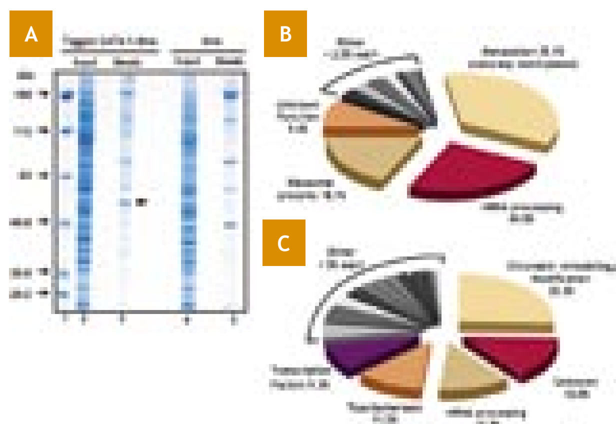


Figure 2 | (A) Colloidal Blue-stained gel of a binding experiment of crude nuclear extracts to streptavidin beads. Lane 1: Marker (M) (see text for further explanation). (B) Classification according to Gene Ontology criteria of proteins identified by mass spectrometry from the control experiment using extracts from cells expressing BirA (around 500 peptide sequences were identified). This represents the background binding. (C) Classification according to Gene Ontology criteria of proteins identified by mass spectrometry using extracts from cells expressing biotin tagged GATA-1 (more than 1000 peptide sequences were identified).

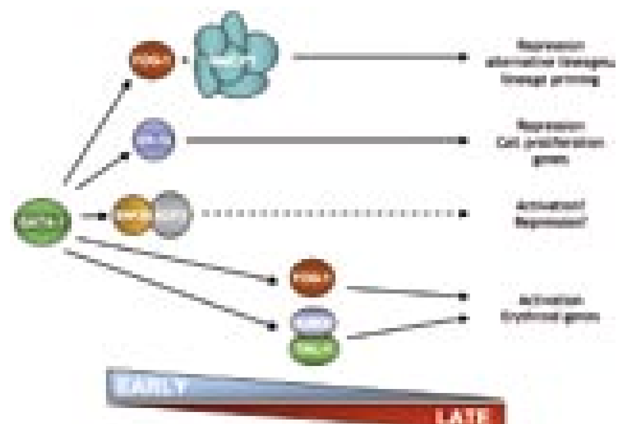


Figure 3 | Model for the distinct GATA-1 complexes and their role in erythropoiesis. Broken arrow indicates unknown function and timing.

the need for extra pre-purification steps. We determined that background in this approach consists primarily of naturally biotinylated proteins, abundant nuclear proteins such as splicing factors, ribosomal proteins and abundant chromatin associated proteins such as topoisomerases, which are purified indirectly together with the tagged transcription factor. This establishes biotinylation tagging as a simple, efficient approach for protein complex purification that is easily accessible to labs with no prior biochemical expertise.An important aspect in hematopoietic development to a particular lineage is the suppression of alternative 'primed' lineage transcription programs and of genes that maintain multipotentiality, while upregulating genes associated with the differentiated cell type. In addition, erythroid terminal differentiation is accompanied by cell cycle arrest. GATA-1 has been implicated in the regulation of all of these aspects.

Our work using biotinylation tagging has identified a number of GATA-1 complexes acting on distinct sets of genes. Thus, we suggest that the GATA-1/Gfi-1b repressive complex acts early in erythroid differentiation and suppresses genes involved in cell proliferation, e.g. myc and myb, while the GATA-1/FOG-1/MeCP1 complex also acts early to suppress genes required to maintain the 'primed' multipotential state, e.g. GATA-2 and alternative hematopoietic lineage genes, e.g. MBP. In contrast, the GATA-1/FOG-1 and the GATA-1/TAL-1/Ldb1 complexes would play a major role in the later upregulation of erythroid genes. This model (see Figure3) is in agreement with recent microarray data examining global expression patterns during erythroid differentiation [8]. Thus, we propose that GATA-1 provides specific early versus late differentiation functions in the context of distinct complexes. The model of different GATA-1 complexes executing specific tasks in different stages of erythroid differentiation suggests a dynamic aspect in the GATA-1 complex interactions during differentiation and also raises the prospect of dissecting the contribution of distinct GATA-1 interactions in erythropoiesis (i.e. essential versus dispensable) by selectively manipulating one specific GATA-1 complex at a time. Lastly, the utility of biotinylation tagging should greatly help in rapidly expanding the characterisation of additional transcription factor complexes, for example, by tagging of the transcription factor partners of GATA-1 and/or of their chromatin co-factors. This in turn will lead to the rapid elucidation of the distinct and overlapping regulatory networks that these factors control in hematopoiesis. Such experiments are currently under way in our laboratory and in collaboration with several other labs.

References

1 Cantor, A.B. and Orkin, S.H. (2002) Transcriptional regulation of erythropoiesis: an affair involving multiple partners. *Oncogene* 21, 3368-3376.

2 Akashi, K. et al. (2003) Transcriptional accessibility for genes of multiple tissues and hematopoietic lineages is hierarchically controlled during early hematopoiesis. *Blood* 101, 383-389.

3 Ferreira, R. at al. (2005) GATA1 function, a paradigm for transcription factors in hematopoiesis. *Mol Cell Biol* 25, 1215-1227.

4 Schatz, P.J. (1993) Use of peptide libraries to map the substrate specificity of a peptide-modifying enzyme: a 13 residue consensus peptide specifies biotinylation in Escherichia coli. *Biotechnology (N Y)* 11, 1138-1143.

5 Beckett, D. et al. (1999) A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Sci* 8, 921-929.

6 de Boer, E. et al.(2003) Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. *Proc Natl Acad Sci U S A* 100, 7480-7485.

7 Rodriguez, P. et al. (2005), J. GATA-1 forms distinct activating and repressive complexes in erythroid cells. *EMBO J* **in press**.

8 Welch, J.J. et al. (2004) Global regulation of erythroid gene expression by transcription factor GATA-1. *Blood*.

John Strouboulis, PhD
Dept. of Cell Biology
Erasmus Medical Center
Dr. Molewaterplein 50
3015 GE Rotterdam
The Netherlands
T +31 10 408 73 52
i.strouboulis@erasmusmc.nl
www.eur.nl/fgg/ch1

Summary

We have described the application of a simple biotinylation tagging approach for the direct purification of tagged transcription factor complexes, based on the use of artificial short peptide tags which are specifically and efficiently biotinylated by the bacterial BirA biotin ligase co-expressed in cells with the tagged factor. We used this approach to initially characterise complexes formed by the hematopoietic transcription factor GATA-1 in erythroid cells. GATA-1 is essential for the erythroid differentiation, its functions encompassing upregulation of erythroid genes, repression of alternative transcription programs and suppression of cell proliferation. However, it was not clear how all of these GATA-1 functions are mediated. Our work describes, for the first time, distinct GATA-1 interactions with the essential hematopoietic factor Gfi-1b, the repressive MeCP1 complex and the chromatin remodelling ACF/WCRF complex, in addition to the known GATA-1/FOG-1 and GATA-1/TAL-1 complexes. We also provided evidence that distinct GATA-1 complexes are associated with specific GATA-1 functions in erythroid differentiation, for example, GATA-1/Gfi-1b with the suppression of cell proliferation and GATA-1/FOG-1/MeCP1 with the repression of other hematopoietic transcription programs. Thus, our work has begun identifying the different protein partners participating in specific GATA-1 functions, has provided novel insight as to the molecular basis of GATA-1 functions, for example through the identification of interactions with histone deacetylases and offers an explanation for the common (and distinct) features of the knock-out phenotypes in erythropoiesis for all of the transcription factors described here.

summary

sander van der krol

cell-specific N-glyco-protein profiling leads to protein markers

Different proteomics technologies are used to analyse complex protein mixtures. In most cases Liquid Chromatography Mass Spectroscopy (LCMS) is used in the final identification step for peptide mass and AA sequence determination. A direct analysis of peptides through LCMS could be most efficient, but this is usually limited by the overall complexity of total protein extracts. One way to lower the complexity is to focus on protein subsets. Constructing transgenic plants in which the endogenous glycoproteins become tagged in specified cells, has proven to allow for the selection of cell-specific glycoprotein pools for protein profiling under different (a)biotic stress signalling events.

The power of '-omics' technologies lies in its unbiased approach to analyse differences between biological samples. Before the '-omics' era scientists were usually trying to confirm a preconceived notion about what is important for a certain process by performing a targeted analysis of their biological samples. In the '-omics' approach we let the experiment decide what may be important and relevant differences between samples, simply by profiling the expression of all genes (transcriptomics), all proteins (proteomics) or all metabolites (metabolomics). In these three branches of the conventional '-omics' techniques the current available protein profiling techniques are least suitable for high throughput (HTP) applications. Although amazing software is available which may 'warp' even the most crooked gels into shape for comparison of protein profiles on 2D-gels, this approach is severely limited by the labour intensive work involved in comparing even a limited set of 2D-profiles. Moreover, the resolution of 2D-gels only allows for the detection of differences between samples of fairly abundant proteins in an extract.

Alternatively, approaches using direct analysis of peptide masses, using different mass spectrometry techniques are under development. However, in general these approaches may face problems in resolution in dealing with the overall complexity of a total protein/peptide mixture. Analysis of protein/peptide pools could be more efficient if the complexity of the protein (peptide) mix that needs to be analysed remains limited, without compromising the potential interest of the selected protein subfraction. Sub-classes of protein pools that may be used for targeted analysis are, for instance, the set of phosphoproteins or the set of ubiquinated proteins that are present in cells. In our case we are targeting the glycoproteins as a subset of the proteome of a cell. Moreover, we are combining the targeting of this sub-class of proteins with a cell specific tagging of N-glycoproteins, thus enabling cell-specific N-glycoprotein profiling.

N-glycan modifications N-Glycosylation is an important post-translational modification of proteins that takes place during

What this research is about:

Biotechnology tool enables cell-specific glycoprotein profiling in plants

The selection of the cell-specific subclass of proteins offers new perspectives for large scale protein analysis because the efficiency of current protein analysis techniques is often limited by the overall complexity of tissue protein extracts. In this article Dr. Sander Van der Krol describes a method for labelling glycoproteins in plants and how, through this label, the tagged glycoprotein pool can be separated from the other proteins in an extract. The labelling can be targeted to specific cell types and this way cell-specifically tagged proteins can be extracted from a whole tissue protein extract.

Glycoproteins are proteins to which different sugars (glycans) are attached to a protein in a cascade of various enzymatic steps. Little is known about the role of glycan modification on glycoproteins in plants, because certain mutants that lack the final steps of glycan modifications are no different from wildtype plants. In contrast, in humans the glycan modifications are crucial for the biological function of glycoproteins. Mammals and plants differ in the end modifications of their glycoproteins. The researchers used this difference to tag N-glycoproteins in plants by introducing a human galactosyltransferase gene in plants, which attaches a galactose sugar residue to the plant glycoproteins. The gene is made active only in a specified cell type, for instance the epidermal cells. A galactose sugar is therefore only attached to plant glycoproteins in epidermal cells. These tagged glycoproteins can be isolated out of the large protein pool for analysis. "The nice thing is that by selection we speed up the analysis," Van der Krol says, "and at the same time, the cell specificity increases the chance of detecting relevant differences. Eventually we want to have a collection of plants with tagging activity in all kinds of interesting cell types. Such a collection of plants will be of interest to many different plant research groups for performing cell-specific glycoprotein profiling within their specific set of experimental conditions."

Research Theme NPC5: High-throughput separation and array-based technologies

β1-4Gal(A)
α1-6Fuc(A)
β1-2Xyl(P)
α1-3Fuc(P)
Glucnac
Mannose
Galactose
Xylose
Fucose

Figure 1 | Core structure of N-glycans and some of the potential modifications occurring in plant (P) and animal (A) systems.

import of these proteins into the Endoplasmatic Reticulum (ER). Upon transfer over the ER membrane a specific glycan (sugar chain structure), which consists of a core antenna of several GlucNacs and Mannose residues, is added to an asparagine (N) residue of the imported protein. The general idea is that the glycan helps protein folding to reach the right conformation. This 'high-mannose' glycan antenna structure is pre-synthesised on the ER membrane and transferred as a whole to the imported protein. After addition to the protein the terminal mannose residues may be trimmed away by different enzymatic activities which reside in the ER and Golgi.

Subsequent modifications of the trimmed down glycan may then occur by addition of different sugar moieties (xylose, fucose, galactose, glucnac, galnac), resulting in so-called 'complex-glycan' structures. Since any of the enzymatic reactions may not go to completion, an enormous heterogeneity is found on glycan structures in eukaryotic cells.

In mammals, the specific composition and structures of the N-glycans are crucial for the biological activity of many proteins and thereby contribute to important cellular processes such as signalling and host-pathogen interactions. Compared to mammals, relatively little is known about glycosylation in plants. The core structures of N-glycans of plants and mammals are similar, but the processing towards complex type glycans may introduce features specific for, for instance, mammals and plants. Figure 1 shows some examples of glycan-core modifications which are typical for mammals and which are typical for plants. It is actually the specific plant-type complex-glycan structure on plant proteins that can provoke such strong immunogenic responses in mammals. The different types of complex glycan structures in mammals and plants indicate that different types of N-glycan modifying enzymes are active in different species.

Glycoprotein tagging In glycoproteomics we make use of the fact that different types of N-glycan modifications occur in different organisms. For instance, in humans a terminal galactose residue is added to the trimmed down N-glycan by the enzyme Galactosyl Transferase (GalT), while this activity is absent in plants. We can therefore use the human GalT gene for the introduction of a heterologous glycan-tagging activity in plants. We have constructed transgenic plants that express the human GalT gene under control of different cell-type specific promoters. The expression of the transgene results in the cell specific production of a novel glycosyltransferase activity, which can tag the glycoprotein pool of the designated cell with a novel sugar moiety (e.g. tagging of glycoproteins in the epidermal cell layer see Figure 2). The introduction of human-type glycan modification into plants has previously been tested by Dr. Bosch's group at Plant Research International, with the purpose of producing functional antibodies in plants with 'human-type' N-glycan modifications. The constitutive expression of the human GalT gene did not lead to a noticeable phenotype in transgenic tobacco plants, while analysis of the N-glycoproteins showed that the endogenous plant glycoprotein pool as well as the expressed antibody were recognised by the introduced GalT activity.

In our case, the novel glycan-tag is used to separate the tagged glycoproteins from all the other proteins in a tissue protein extract. The tagged proteins can be isolated by affinity chromatography, using either lectins (e.g. RCA lectin for the galactose-tagged plant glycoproteins) or antibodies that specifically recognise the novel glycan-tag. Alternatively to purifying the tagged intact glycoprotein, the total protein pool

may first be treated with trypsin to generate specific peptide fragments. From this pool the glycan-tagged peptides may be purified, so that only one peptide per protein needs to be analysed. In the last step, the glycan is removed from the glycopeptides with PNGase treatment, after which the mass of the peptide is determined by LCMS. The cell specific tagged-glycopeptide pool is up to 1000-fold reduced in complexity compared to that of a peptide mix from a full protein tissue extract. When purification steps are optimised, there may therefore be a true potential for HTP handling of samples using direct LCMS.

Why Glyco-proteomics? Selecting a sub-pool of proteins for profiling should not just be a technical opportunistic chance for performing effective proteomics, but should also have a biological relevance. In that respect the selection of glycoproteins in plants makes an interesting sub-class of the proteome, as very little is yet known about the biological function of complex N-glycosylation in plants. Using different bioinformatics tools to analyse the *Arabidopsis* genome sequence, we have determined the potential glycoprotein pool of *Arabidopsis*. The analysis shows that approximately 5000 of the 30,000 proteins of the *Arabidopsis* proteome contain a peptide signal sequence indicative of processing through the ER and Golgi. Of this group of proteins approximately 4000 proteins contain one or more of the potential N-glycosylation AA-consensus recognition sequences (N-X-S/T). Because not every N-glycosylation site on a protein is actually glycosylated, this number thus forms the upper limit of potential glycoproteins in plants. Unlike in animals, where defects in N-glycosylation are associated with severe disease symptoms, the role of N-glycosylation in plants is far less clear. Only 20% of these putative glycoproteins have been functionally annotated, indicating that very little is yet known about their biological role in plants. More than 75% of the putative glycoprotein pool of plants consists of proteins containing one or more membrane-spanning domain, indicative of a function at the interface of sub cellular compartments or at the interface of the cell and its environment. Cell specific glycoproteome profiling is therefore ideally suited for investigating signalling events during biotic (caused by e.g. pathogens, interacting bacteria or fungi) and abiotic (caused by e.g. light, temperature or mechanical stress) signalling in plants.

Plant pathogen markers As the first interaction of many plant pathogens is at the site of the epidermal cell layer, we are focussing our attention on epidermal specific glycoprotein tagging, both in the model species *Arabidopsi*s as well as in the important crop species tomato and potato. Transgenic plants, which contain the human GalT gene under epidermal specific promoter, have been generated. When we want to select tagged epidermal proteins from the infected plant tissue we have to be sure that the introduced glycan-tag is unique both for the plant as well as for the pathogen. For this purpose we have confirmed that the glycan structure of glycoproteins from the plant pathogens we want to study is compatible with our tagging strategy in plants. The tagging strategy will also be applied to the plant pathogen, allowing for the selection of pathogen tagged glycoproteins from an infected leaf extract. With the specificity of cell specific proteomics we thus hope to detect the very early response proteins in both plant and pathogen.

Eventually a collection of plants with different types of cell specific glycoprotein tagging activities (epidermis, mesophyll, phloem, different meristematic regions) can be used for cell specific protein profiling to study plant development or different biotic or abiotic stress signalling events. Identification of specific differences in the profiles will help identify protein markers, which may be used to diagnose for these infections or which may be used in selective breeding. Moreover, the inventory of glycoprotein profiles will certainly aid in a better understanding of the glycosylation in plants and may be of help in the functional annotation of the large group of unknown proteins that are part of the glycoprotein sub-proteome.

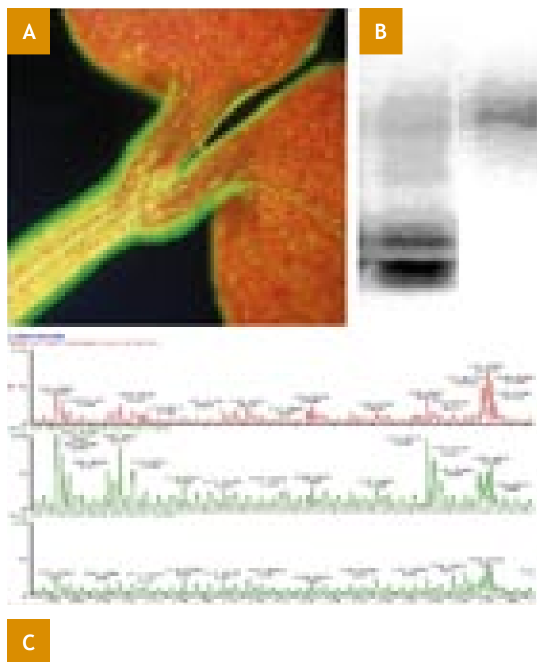


Figure 2 | A. In the cell layer, which is marked by the green fluorescence, a human GalT gene for cell specific modifications of glycoproteins is expressed. B. Western blot of wt *Arabidopsis* and transgenic arabidopsis with epidermal specific GalT activity. Glycoproteins with glycans that are modified with a terminal galactose are detected with RCA lectin. In wildtype plants there is only one band visible (right), while in the protein extract from the epidermal glycan-tagged plant multiple protein bands light up (left). C. MS profiles of peptide fractions collected during glycan-affinity chromatography. top: input peptide fraction, middle: upbound peptide fraction, bottom: bound peptide fraction.

Sander van der Krol, PhD
Plant Physiology
Laboratory of Plant Physiology
Department of Plant Science
Wageningen University
Arboretumlaan 4
6703 BD Wageningen
The Netherlands
T +31 317 482 448
sander.vanderkrol@wur.nl

Summary

Current available protein profiling techniques (proteomics) are not very suitable for high-throughput (HTP) applications. If the complexity of the protein (peptide) mix is reduced, a direct analysis by LCMS could become more efficient. In our research we obtain this by focussing on the subpool of N-glycosylated proteins in plants. Glycoproteins pass through the export system of cells and many end up in the cell membrane or the extracellular space. For this reason they are of particular interest in cell-cell or cell-environment interaction studies. The biological function of more than half of the glycoproteins is unknown.

N-Glycosylation is an important post-translational modification of proteins that takes place in the Endoplasmatic Reticulum (ER) and Golgi. A specific glycan (a core of glucnacs with several mannose residues) is added to the asparagine (N) residue of the protein (hence the name N-glycosylation). There are specific different types of modifications of the core structure of glycans that are executed by enzymes specific for different species.

We made use of this difference between animal and plants by constructing a transgenic plant containing a human galactosyl transferase (GalT) gene. By placing this human GalT gene under the control of different cell specific promoters, we are able to construct plants which selectively modify the glycans in specified cell types (e.g. epidermal cells). The glycan modification serves as a tag, which allows isolation of the tagged glycoprotein pool from a specified cell type out of a total protein extract. Such N-glycoprotein profiling of specified cell types will be put to use first in plant-pathogen interaction studies.

summary

ron heeren

mass spectrometer identifies and localises biomarkers in cells

When Anthony van Leeuwenhoek first looked through his microscope, he could not have envisaged how his eye as detector would come to be replaced by fast, sensitive cameras and his illuminating candle by intense light sources and lasers. Now, microspectroscopic imaging tools are used to study the distribution and interactions of biomolecules within cells and tissue.

Chemical imaging mass spectrometry is revolutionising the field of biological surface analysis. It provides both the chemical information of a mass spectrometer and the spatial organisation of each component on a surface. High spatial resolution combined with high sensitivity for large molecules remains the goal of chemical imaging mass spectrometry. Both matrix-assisted laser desorption/ionisation (MALDI) and secondary ion mass spectrometry (SIMS) based approaches are being developed to achieve this target. These two techniques offer different starting positions for the ultimate goal of high mass, high-resolution imaging, namely high mass and relatively low spatial resolution (MALDI) and low mass and high spatial resolution (SIMS).

The research within the NPC aims at identifying and localising biomarkers in cells and tissues using MS imaging techniques. Knowledge of particular biomarkers may be used to diagnose illnesses or abnormalities such as tumours, it may be used to develop a therapy or the pathology of the illness may be studied.

Through the use of well-established multiple fluorescent labelling techniques (colouring of tissue), it is possible to study the spatiotemporal behaviour of selected biomolecules in parallel in a single experiment. Despite the valuable insights already realised, any labelling technique suffers from several innate weaknesses. Arguably the most important is the need to identify the molecules of interest prior to the experiment in order to apply the appropriate label. In addition, the small number of molecules that can be investigated simultaneously in a single experiment, the possible interference of the label with the protein's normal function, and the difficulties in distinguishing between specific posttranslational modifications of proteins limit the application of labelling techniques as a discovery tool.

This last aspect is especially important considering that several diseases and immune responses have been associated with altered posttranslational modifications. Clearly, the desire to study disease and the complexity of biological mechanisms demand a molecule-specific imaging technique that can be applied to tissues or cells directly. Mass spectrometry possesses the chemical specificity and high sensitivity required for biomolecular imaging experiments without the use of chemical modifications or labels.

Two main approaches were taken towards chemical imaging mass spectrometry at AMOLF. Firstly, matrix-enhanced SIMS

What this research is about:

Exploiting physical knowledge for biological purposes

"There is no single method that combines biomolecular sensitivity with high spatial resolution," says prof. dr. Ron Heeren. He aims at chemical imaging mass spectrometry, which he developed to look at specific biomarkers in cells and tissue. Heeren's group is specialised in the physics of mass spectrometry and has been involved in imaging since 1995. He defines his activities as 'exploiting physical knowledge for biological purposes'.

Imaging mass spectrometry provides both chemical information and the spatial distribution of each analyte detected. A challenge in this field is to combine high spatial resolution with high sensitivity for high-mass molecules. Heeren and his team used two methods to this effect. Matrix-enhanced SIMS imaging is used to gather chemico-spatial information that cannot be provided by established SIMS or MALDI imaging techniques. The addition of a matrix to a sample improves the sensitivity of SIMS for higher mass molecules. When combined with a primary ion column capable of delivering high-resolution images, such as a liquid metal ion gun or an ion-optical microscope, ME-SIMS allows images of intact molecular ions to be recorded with high spatial resolution.

Secondly, Heeren used a novel approach to MALDI molecular imaging that combines high spatial resolution with high-speed analysis while operated at the preferable lower laser fluences. The group has demonstrated the feasibility of MALDI microscope mode imaging. The test samples that were used delivered the recognisable structures needed to prove the concept of MALDI microscope mode imaging. In addition, they served as useful targets to estimate the experimental spatial resolution. The high spatial resolution, the increased speed and the greater versatility of potential ion sources offered by the mass microscope approach will prove important in the development of the mass spectral molecular imaging of biomolecules.

Research Theme NPC6: Biomarker discovery

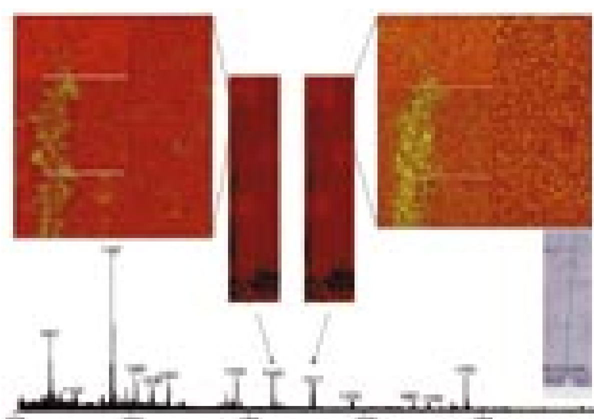


Figure 1 | **ME-SIMS images of rat brain tissue.** The biomarker at *m/z* = 1511 is present in all cells, the biomarker at *m/z* = 1443 is localised in specific cells.

(ME-SIMS) was applied to mollusc, insect and rat brain tissue. The high spatial resolution images, in the low micrometer range, of intact molecular ions demonstrate the ability of ME-SIMS imaging to provide chemico-spatial information that cannot be provided by established SIMS or MALDI imaging techniques.

Secondly, a novel MALDI microprobe instrument has recently been reported that is capable of imaging mass spectrometry with a micrometer-size laser probe. The drawback of this approach is the high laser fluence needed to reach the ionization threshold for MALDI from such a small spot, which leads to a high degree of fragmentation and a significant loss of sensitivity. Moreover, because each pixel is sequentially analysed, to increase the spatial resolution by a factor *N* requires an increase in analysis time of *N*2. We present a novel approach to MALDI molecular imaging that combines high spatial resolution with high-speed analysis while operated at the preferable lower laser fluences.

Brains in a matrix In ME-SIMS, a sample is prepared in a matrix which increases sensitivity for high-mass molecules with respect to conventional SIMS. Analysis of cockroach brain tissue shows that the ME-SIMS images clearly resolve features that are separated by less than 10 μm, and that the cholesterol, choline and diacylglycerol distributions differ significantly.

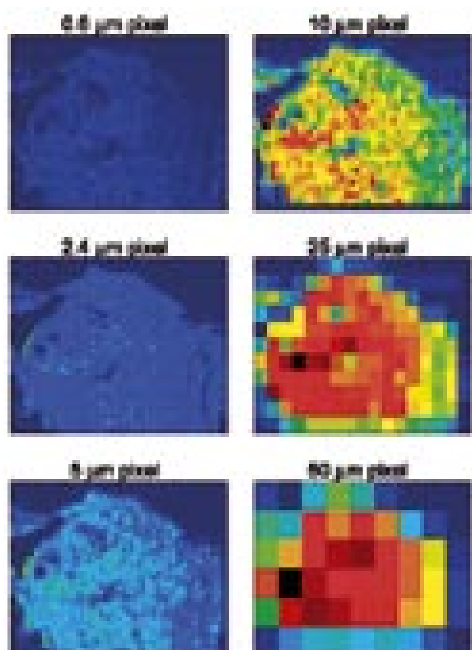


Figure 2 | The effect of resolution on images of cholesterol in tissue by ME-SIMS.

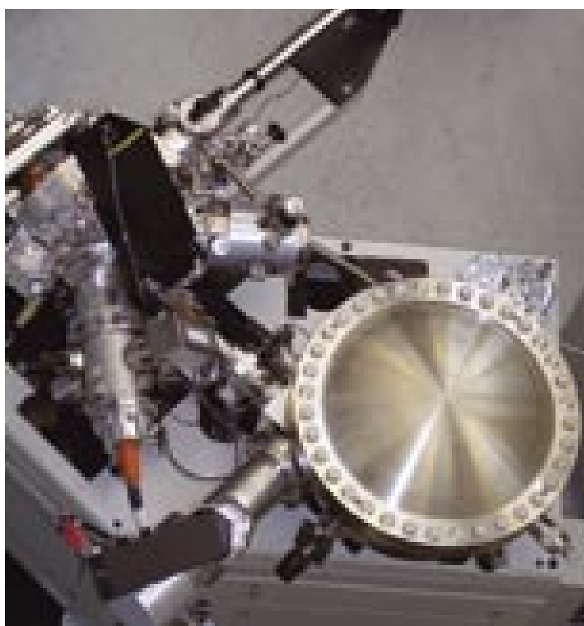


Figure 3 | Imaging SIMS set-up.

To obtain information from within an organ, thin tissue sections can be cut using a microtome. The enhanced information obtained from tissue sections was demonstrated with SIMS spectra obtained from a thin section of the parietal ganglion of Lymnaea stagnalis before and after matrix deposition.

The results demonstrate that ME-SIMS provides both the spatial resolution and the sensitivity for molecular ions necessary to reveal subcellular distributions of analytes, in addition to providing an in situ quality assessment of the images obtained (height map). Without matrix the low yields and lack of chemical specificity of the resulting low-mass fragments prompted a leading figure in the SIMS community to reflect that, 'There are few compounds that exist at sufficient concentrations and display sufficient secondary ion yields to permit any sensible spatial analysis' [1]. The addition of a matrix helps to alleviate these sensitivity/specificity problems. This facile improvement on the technique provides information not currently obtainable with regular SIMS or conventional ≥25 µm laser spot MALDI chemical imaging mass spectrometers.

Finally, rat brain tissue was prepared using gold deposition and analysed using ME-SIMS. Images were obtained using a microprobe with a 100 nm beam, giving images of 150x150 µm. It was clearly shown that one of the biomarkers with *m/z* 1443 was localised to specific cells, whereas another of *m/z* 1511 was present in all cells in the brain tissue. There is no other method which combines this biomolecular sensitivity with high spatial resolution.

Microscope One disadvantage of ME-SIMS combined with the mass microscope used at our laboratory is the limitation of molecule size up to ~2000 Dalton. Using MALDI on the same instrument, molecules up to 10,000 Dalton can be imaged. We have used an ion microscope in combination with MALDI-MS to record macromolecular images of the spatial distributions of intact peptide and protein ions [2]. Using a mass spectrometric microscope, the spatial resolution is independent of the spot size of the ionizing beam. This decoupling allows a much larger area to be examined without having to move the sample or the laser spot. The ions produced by a single laser shot pass through the time-of-flight mass spectrometer forming an ion-optical image on a position sensitive detector, much like in wide-field optical microscopy. In this manner, a mass-to-charge ratio (*m/z*) separated series of molecular images is generated showing spatial detail from within the laser spot.

The small features obtained in our MALDI microscope experiments would not have been discernible with the ≥25 µm spot sizes of a typical MALDI microprobe. In contrast, multiple squares identified in the images would have been sampled by a single laser position and thus be represented as a single pixel. The spatial resolution of around 4 µm obtained in this current work is comparable to that of the micrometer-size microprobe studies. Moreover, these high-resolution images were recorded without incurring the loss of sensitivity at increased laser threshold fluences associated with a highly focussed laser.

A distinct advantage of the microscope over the microprobe approach is its improved speed of analysis. Compared to a 2 µm microprobe (at least two pixels are necessary for 4 µm experimental resolution), the mass microscope can analyse the same area, delivering equivalent performance, using much fewer individual positions. In addition, the decoupling of the spatial resolution from the source conditions offers an advantageous extension of useful ionization methodologies for imaging purposes. In particular, the microscope approach allows the use of ion sources that cannot provide convenient small spot sizes for microprobe experiments, such as infrared laser sources in IR-MALDI.

Time for validation We can conclude that the addition of a matrix to a sample improves the sensitivity of SIMS for higher mass molecules. When combined with a primary ion column capable of delivering high-resolution images, such as a liquid

metal ion gun, or an ion-optical microscope, ME-SIMS allows images of intact molecular ions to be recorded with high spatial resolution. The mechanistic differences between ME-SIMS and MALDI suggest that there is significant scope for improving the sensitivity of ME-SIMS for macromolecules. This would improve the contrast of the images already attainable in addition to increasing the number of analytes amenable to high spatial resolution analysis.

Furthermore, we have demonstrated the feasibility of MALDI microscope mode imaging. The test samples we have used delivered us the recognisable structures needed to prove the concept of MALDI microscope mode imaging. In addition, they served as useful targets to estimate the experimental spatial resolution. The high spatial resolution, the increased speed, and the greater versatility of potential ion sources offered by the mass microscope approach will prove important in the development of the mass spectral molecular imaging of biomolecules. The quality of the results obtained justifies further studies in which microscope mode imaging will be applied to biological samples.

Using chemical imaging mass spectrometry we have taken a clear step towards localising biomarkers that are present in tissue only in very small amounts. Their identification will rely heavily on the NPC infrastructure provided in the biomarker discovery program. Building infrastructure, initiating cooperations and gathering expertise have been our main occupation. Now it is time to validate and optimise our methods for application in diagnosis and pathology.

References

1 McDonnell, L.A. et al. (2005) Subcellular imaging mass spectrometry of brain tissue. *J. Mass Spectrom.* 40, 160–168.
2 Luxembourg, S.L. et al. (2004) High-Spatial Resolution Mass Spectrometric Imaging of Peptide and Protein Distributions on a Surface. *Anal. Chem.* 76, 5339-5344.

Ron Heeren, PhD, Prof.
FOM Institute for Atomic and Molecular Physics
Kruislaan 407
1098 SJ Amsterdam
T +31 20 608 12 34
R.Heeren@amolf.nl
www.amolf.nl/research/macromolecular_ion_physics/main.html

Summary

Chemical imaging mass spectrometry provides both the chemical information of a mass spectrometer and the spatial organisation of each component on a surface. The research within the NPC aims at identifying and localising biomarkers in cells and tissue using MS imaging techniques. Knowledge of particular biomarkers may be used to diagnose illnesses and abnormalities such as tumours, it may be used to develop a therapy or the pathology of the illness may be studied.

Through the use of well-established multiple fluorescent labelling techniques (colouring of tissue), it is possible to study the spatiotemporal behaviour of biomolecules in parallel in a single experiment. The disadvantages of this technique include the need to identify the molecules of interest prior to the experiment, and the small number of molecules that can be investigated simultaneously. Clearly, the desire to study disease and the complexity of biological mechanisms demand a molecule-specific imaging technique that can be applied to tissues or cells directly. Mass spectrometry possesses the chemical specificity and high sensitivity required for biomolecular imaging experiments without the use of chemical modifications or labels.

Two main approaches were taken towards chemical imaging mass spectrometry. Firstly, matrix-enhanced SIMS (ME-SIMS) was applied to mollusc, insect and rat brain tissue. High spatial resolution images were obtained and it was demonstrated that ME-SIMS has the ability to provide chemico-spatial information that cannot be provided by established SIMS or MALDI imaging techniques. Secondly, a novel approach to MALDI molecular imaging is presented that combines high spatial resolution with high-speed analysis while operated at the preferable lower laser fluences. Using these techniques, we have taken a step towards finding biomarkers that are present in tissue only in very small amounts.

summary

upcoming events

28 August - 1 September 2005 | HUPO 4th Annual World Congress: From defining the proteome to understanding function, Munich, Germany

5 October 2005 | Genomics Momentum, The Hague, The Netherlands

4 November 2005 | NPC Theme 4 Symposium: Regulatory protein networks, Utrecht, The Netherlands

2 February 2006 | NPC Midterm Progress Meeting, Utrecht, The Netherlands

perspective on proteomics

genome canada

Cindy L. Bell, Ph.D.
V.P. National
Genomics Program
Genome Canada

Chuck Hasel, Ph.D., M.B.A
Scientific Officer
Genome Canada

Since its inception in 2003, the Netherlands Proteomics Centre (NPC) has made significant headway towards its goal of fostering outstanding proteomics research in the Netherlands and should be commended for conversion of a concept into reality in such a short period of time.

The NPC has been instrumental in promoting partnerships between universities, medical centres and the biotech industry. As proteomics research is inherently complex in nature and rapidly evolving, the NPC concept and implementation of the research hotels is brilliant, as it allows novice researchers to learn from experts over an extended period of time.

The Centre's goal of improving and widening fundamental technological tools in proteomics research is important and the NPC should be applauded for making technology development and training a key component of the Centre. In addition, the NPC program is clearly attacking some topical research areas, such as biomarkers and membrane proteomics, which are highly relevant to advancing human health research.

The success of the NPC is further demonstrated by the number of publications, the ability to attract trainees, an emerging patent portfolio and the contribution to a start-up company.

To date, the NPC has focused on support of the national research community, and done so very well. It will be interesting and important to extend that focus in the coming years to the European research community, as well as, towards further international collaborations.

The NPC has developed an excellent model, which will be informative for Genome Canada, as it designs and implements a strategy for proteomics research. As evidenced by its successful launch, the NPC should continue to meet and surpass its future objectives on its way to becoming an internationally recognized organization.