Identification of nodule-specific NCR peptides, comparative proteome analysis of wild type and mutant *Medicago truncatula* nodules and identification of potential interacting protein targets of NCR247 in endosymbiotic bacteria

Thesis of the PhD dissertation

Hajnalka Dürgő

Supervisor: Dr. Katalin F. Medzihradszky

PhD School in Biology, University of Szeged, Faculty of Science and Informatics

HAS Biological Research Centre, Institute of Biochemistry

Szeged

2015

1. Introduction

Rhizobium bacteria are Gram-negative bacteria that live in a symbiotic relationship with host plants. The root cause of the symbiosis is the low level of NH_4^+ in the soil. The bacteria satisfy the nitrogen need of the host by reducing the atmospheric nitrogen to ammonia, while the plant provides carbon sources and energy for the bacteria. This plant-bacterium interaction is host-specific and molecular communication between the partners leads to the formation of a symbiotic organ, the root nodule. In the nodules, the bacteria adapt to the intracellular life and differentiate progressively to nitrogen-fixing bacteroids. In *Medicago truncatula* nodules, the differentiation of the *Sinorhizobium meliloti* or *S. medicae* bacterial partner is irreversible leading to the formation of large polyploid non-cultivable bacteroids.

There is a continuous interaction between the plant and the bacteria during the nodule formation. One of the key steps of this process is the transcription of plant genes dormant before the inoculation. The participating genes are classified as early (Enod) or late (Late-nod) nodulin genes depending on whether they are active at the beginning of the plant-bacterium interaction or after nodule formation, respectively.

The plant effectors are the nodule-specific symbiotic peptides (*symPEP*). In the *M. truncatula* genome, at least 560 small genes code for such peptides. Expression of these genes exhibits extreme nodule-specificity and requires Rhizobium infection. The transcripts of the different *symPEP* genes can be

localized in different nodule zones representing various stages of symbiotic cell maturation, and they are produced in a large amount in the nodules. *SymPEPs* can be classified by their composition into 3 groups. We focused on the nodule-specific cysteine-rich peptides (NCRs, the largest group). More than 500 genes code for NCR peptides in *Medicago truncatula* nodules. These polypeptides contain a signal peptide and a mostly 30-50 residue-long mature chain. The secreted mature peptides feature conserved cysteine patterns (4 or 6 Cys). Due to their highly variable amino acid compositions their pI's range from 3.2-11.25. NCRs have unique amino acid sequences, but because of their size and cysteine-content they resemble defensins, the largest group of antimicrobial peptides in plants. The NCR's toxic effects seem to be linked to their cationic or anionic character. The antibiotic activity against some Gram-negative and Gram-positive bacteria and human pathogenic fungi was confirmed in several publications (*in-vitro* studies).

Cationic NCR peptides target the bacterial membrane, anionic peptides accumulate in the bacterial cytosol. As a result of the NCR's action the cell membrane becomes permeable. Bacterial respiration and reproduction will be compromised. The potential intracellular targets of the NCR peptides might be proteins involved in bacterial proliferation.

2. Aims

In this study we focused on the detection of the most important plant protein group in the nitrogen-fixing symbiosis: the NCR peptides. Despite of genomic data and extensive transcriptome studies until now the existence of NCRs have not been proven at the peptide level. Many questions were unanswered: how many NCR products exist, could we detect them, when they do appear, what their turnover rate is, which of these peptides interact with and accumulate in the endosymbionts. *M. truncatula* A17 Jemalong wild type plants were inoculated with *Sinorhizobium medicae* and *S. meliloti* bacteria, experiments were performed with the indeterminant nodules of these plants. We analyzed the nodule-proteome of both the wild type and the 6V mutant plant that is unable of nitrogen-fixing.

Our aims were as follows:

1) The first aim of our examination was to detect these plant peptides which regulate bacterial differentiation in bacteroids. Additionally we intended to validate signal peptide predictions, i.e. to determine the signal peptidase enzyme's processing sites in these NCRs.

2) We tried to decipher the biological functions of these peptides.

    a) We attempted to determine differences between the NCR-contents in bacteroid populations at different developmental stages.

b) We compared the NCR- and protein-content of bacteroids isolated from wild type and mutant plant nodules. The selected 6V/*dnf* mutant has a well-defined phenotype.

c) We investigated the bacterial and plant protein targets of an NCR peptide with verified antimicrobial effect.

3. Methods

1. Different bacteroid purification methods: i) for a bacteroid mixture representing all developmental stages ("short" purification) and ii) for fully differentiated bacteroids (Percoll-gradient purification)
2. Bacteroid disruption with cyclic sonication
3. Affinity chromatography of bacteroid extracts using StrepII or Flag-tagged NCR247
4. Protein digestion (in-gel and in-solution digestions)
5. Protein identification based on the LC-MS/MS analysis of the digests, and database searching (with Proteome Prospector software, in Uniprot and an "in-house" NCR database)
6. Comparison of the NCR-content of bacteroid samples representing different developmental stages using the fitXIC software
7. Semi-quantitative comparison of the bacteroid proteome present in wild-type and a mutant plant nodule (spectral counting)

4. Results

1. As the primary aim of this study, we focused on large-scale detection of NCR peptides with mass spectrometry. Three types of biological material were studied, based on the following assumptions. The root nodules are in principle the best suited for the identification of all expressed NCRs. However, the plant and bacterial protein background in nodules might hinder the detection of NCR peptides. Thus, the bacteroids were isolated using two purification protocols. The short one resulted in the isolation of a mixture of endosymbionts in all developmental stages, whereas the long one utilizing a Percoll gradient centrifugation enriched the terminally differentiated nitrogen-fixing bacteroids. From the nodules 2, and from the bacteroid samples 3 biological replicates were prepared.

Almost 300 proteins were identified from the combined data acquired from the intact nodule samples. More than half of these proteins derived from the plant, including 45 NCRs.

The analysis of protein content of bacteroids in all developmental stages ('short purification') yielded approximately 600 proteins; more than 2/3 of these originated from the bacteria, and 118 NCR peptides were detected. Analysis of the 'mature' nitrogen-fixing bacteroids (Percoll-gradient purification) yielded 313 bacterial proteins and 103 *M. truncatula* proteins, including 75 NCR peptides.

Altogether, 138 NCRs were detected in these experiments. The results show that the further isolation of the intact bacterial sample is necessary for the deeper protein analysis.

In several NCRs we successfully identified the N-terminus of the mature peptide, thus, the cleavage site of the signal peptidase was confirmed.

In order to identify potential differences in the NCR-content of the different bacteroid samples we performed an MS-based data comparison. A reference list was created from the merged search results, listing only the best CID data for each NCR sequence with the precursor's m/z and the retention time. The data were compared to this reference list using the fitXIC program. An NCR entry was considered specific to a preparation if at least one tryptic peptide was detected in two of the biological replicates, while none was detected in the other preparation.

Twelve NCR peptides were found only in the 'short purification' mixture. We believe these peptides may be involved in the early stages of the bacteroid development. The 'Percoll purification' featured 5 unique NCR peptides. The persistence of these peptides suggests later or long term functional involvement in the endosymbionts.

2. Next we compared the proteome of wild type and 6V (*dnf7*) mutant nodules (inoculated with *S. medicae*). This mutant was selected because this plant produces aberrant nodules in which bacteria are unable to fix nitrogen. The NCR169 gene is absent in this mutant, and this is the reason for the above mentioned defect. Thus, the changes in the proteome of the mutant nodules reflect the influence of this particular NCR. The protein contents of bacteroid mixtures were compared, 3 biological replicates, with 3 technical replicates each. The workflow introduced in the earlier study was followed

except for one sample we also performed protein-level fractionation on SDS-PAGE prior to the tryptic digestion and MS analysis. Quantitative data analysis was performed using spectral counting. A 50% change in the protein levels was considered significant. In mutant nodules, 16 proteins were present at higher levels, their significance requires further studies. Twenty three proteins were 'lost' or decreased in mutant samples. Series of proteins participating in nitrogen-fixing process, such as NifT/FixU family protein, Ferredoxin III 4(4Fe-4S) nif-specific, Nitrogenase protein, Nitrogen fixation protein NifX were detected at a significantly lower level than in the wild-type plant, thus our results definitely reveal the defect of nitrogen-fixation on the proteome level in the *M. truncatula* 6V mutant. In addition to these changes several NCR and GRP peptides (*symPEP*s) were detected exclusively in the wild type nodules. The in-solution and in-gel digestions produced slightly different results. The in-gel protein-fractionation was more suitable for larger proteins, but the majority of smaller NCR peptides were identified from the in-solution digested unfractionated samples. It indicates that these two approaches could supplement each other.

Altogether, 139 NCR peptides were detected in this experiment. Almost half of them were found only in the wild type nodules and none of them was unique for the mutant. Some NCR peptides were detected the first time.

3. Eventually we tried to decipher the biological role of NCR247 by identification of its potential interacting protein partners. This NCR was selected, because it showed antibiotic activity against some Gram-negative, Gram-positive pathogenic bacteria. Transcription analysis of NCR247 shows peptide production in the early stage of nodule development. Chemically

synthetized NCR247 (without disulfide bonds) was used as bait for the protein-interaction analysis. The peptide was used with StrepII- and FLAG-tag for the easier isolation of the protein complexes. *S. meliloti* bacterial culture and bacteroids (from *M. truncatula* nodules) were studied (7 bacterial and one bacteroid sample was analyzed in parallel with control experiments, *in vitro* experiments).

The potential targets in bacterial samples are: ribosomal proteins (14 and 12 proteins from the small and large subunits); bacterial chaperonin GroEL, pyruvate dehydrogenase complex, transaldolase, RNA polymerase subunit beta and beta', elongation factors and some other proteins. The potential interacting partners from bacterial and bacteroid samples were grouped into functional categories. From free-living bacteria samples the most abundant interacting proteins were the ribosomal and the nucleotid-binding proteins.

In the bacteroid sample, GroEL, subunits of pyruvate dehydrogenase complex and ribosomal proteins were detected as interacting partners. Additionally, units of the nitrogenase complex and NCR peptides (NCR028, NCR169, NCR290) were identified. The amount of the detected protein targets decreased in bacteroid samples implicitly, the most abundant groups were: NCR peptides, transport proteins, ribosomal and nitrogen-fixing protein groups.

The relevance of NCR247-GroEL relationship was confirmed by immuno-precipitation. Presently it is not clear whether NCR247 directly interacts with proteins from the nitrogen-fixing pathway and with other NCR peptides or

through intermediates. The significance of these interactions should be further investigated.

5. Summary

1. Our work represents the first large scale demonstration of NCR peptides, confirming translation of the NCR genes and high level accumulation of the NCR peptides in the bacteroids. Almost 200 NCR peptides were identified with mass spectrometry. In several NCRs we successfully identified the N-terminus of the mature peptide, thus, the cleavage site of the signal peptidase was identified. Comparison of NCR contents in nodules in the early stages and the fully differentiated bacteroids was also performed. . We identified some unique NCR peptides in the early nodule-development stages.
2. We performed a semi-quantitative comparison of the proteome of wild type and 6V (dnf7) mutant (unable of nitrogen-fixing) *M. truncatula* nodules. Our results show that the production of nitrogen-fixing proteins is reduced in the mutant plant in the absence of NCR169. Some NCR peptides were identified only in the wild-type samples suggesting that their expression is dependent on prior actions controlled by NCR169.
3. Potential protein targets of NCR247 (NCR with antimicrobial effect) were identified in bacterial and bacteroid samples. Ribosomal proteins, GroEl chaperon protein and other proteins including NCR peptides show affinity to NCR247.

4. Publications

Publications related to this thesis:

Attila Farkas, Gergely Maróti, **Hajnalka Dürgo**, Zoltán Györgypál, Rui M. Lima, Katalin F. Medzihradszky, Attila Kereszt, Peter Mergaert, and Éva Kondorosi. The Medicago truncatula symbiotic peptide NCR27 contributes to bacteroid differentiation through multiple mechanisms. *Proc Natl Acad Sci U S A* **2014,** *111*, 5183-5188.

IF: 9.809

Hajnalka Durgo, Eva Klement, Eva Hunyadi-Gulyas, Attila Szucs, Attila Kereszt, Katalin F. Medzihradszky, Eva Kondorosi. Identification of Nodule-Specific Cysteine-Rich Plant Peptides in Endosymbiotic Bacteria.

Proteomics, *in press*

IF: 3.973

Other publication:

Fekete A, Kenesi E, Hunyadi-Gulyas E, **Durgo H**, Berko B, Dunai ZA, Bauer PI. The guanine-quadruplex structure in the human c-myc gene's promoter is converted into B-DNA form by the human poly(ADP-ribose)polymerase-1. *PLoS One*. **2012**, *7,* e42690. doi: 10.1371PLoS One, IF: 4.092

Σ IF: 17.874