

**Structural genomic and transcriptome analysis of the
sulfanilic acid degrading *Novosphingobium resinovorum*
SA1 strain**

Ph.D. Thesis

Botond Hegedüs

Supervisor:

Dr. Gábor Rákhely

Doctoral School of Biology

University of Szeged, Faculty of Science and Informatics, Department of
Biotechnology, Hungarian Academy of Sciences, Biological Research Center

Institute of Biophysics

Szeged

2018

INTRODUCTION

In order to overcome current challenges of human society, industries must go through continuous developments. As a consequence, raw materials derived from natural sources are usually replaced by chemically modified or fully synthetic compounds. These materials are produced in large quantities in standard quality and they can be endowed with properties which cannot be found in nature. However, during their production and usage they are inevitably released into the environment where - due to their xenobiotic characteristics - their natural degradation is very slow or does not take place at all. This can lead to their local accumulation, or in a worse case, to their incorporation into the food chain. Sulfanilic acid (SA) is one of these artificial compounds which is a widely used and manufactured key intermediate in production of azo dyes, plant protectives and pharmaceuticals. Its natural degradation - due to its aromatic structure and unnatural sulfonate group - is very slow. Despite of its recalcitrant nature, a few bacterial strains became capable of its degradation and usage as nutrient. One of these strains is the *Novosphingobium resinovorum* SA1 belonging to the *Novosphingobium* genus. The strains of this genus usually have diverse metabolic pathways and highly efficient adaptability. These properties made it a promising candidate to degrade and utilize the SA as a nutrient source. During my research, my aim was to determine the whole genome sequence of the *N. resinovorum* SA1 strain, in order to reveal the full structure of the SA metabolic pathway and its evolutionary origin. Additionally, I used RNA-Seq method for transcriptome of *N. resinovorum* SA1 cells grown in glucose and sulfanilic acid with samples being taken from the exponential and stationary growth phases in order to assess global effect of SA and the likely associated starvation on the gene expression. In addition to the SA catabolic genes I tried to identify the components of the auxiliary systems which play an indirect but important roles in the effective degradation process.

METHODS

During the sequencing project, three different genome libraries were sequenced. In all cases, Illumina MiSeq sequencing platform was used. Paired-end and mate paired-end reads obtained during sequencing were combined in a 'hybrid assembly' approach.

The remaining uncertain or incomplete parts, gaps were clarified by means of polymerase chain reaction (PCR) and capillary sequencing.

The annotation of the genome was carried out with the NCBI (National Center for Biotechnology Information) PGAAP (Prokaryotic Genome Autonomic Annotation Pipeline) pipeline and became available in the NCBI GenBank as well as in the Ensemble Genomes databases.

The annotation was also carried out using the UniProt (Universal Protein Resource) and KEGG (Kyoto Encyclopedia of Genes and Genomes) resources and is available in their databases, as well. Protein classification into functional categories of COG (Clusters of Orthologous Groups of proteins) was carried out with the eggNOG-mapper software package based on the eggNOG 4.5.1 database.

Global (Needleman-Wunsch) alignment was used to measure the sequence identities between the SA catabolic pathway elements identified from different strains.

Identification of the possible homologs of the SA catabolic pathway elements in the NCBI „RefSeq Representative genomes” database was performed with the NCBI TBLASTN software package.

The multiple sequence alignment required for the phylogenetic analysis was performed with the MAFFT 7 software package applying the E-INS-I iterative run mode. The phylogenetic tree was calculated with the MEGA 6 program using the maximum likelihood method.

The relative synonymous codon usage (RSCU) of the *N. resinovorum* SA1 protein coding genes were analyzed with correspondence analysis (CA).

In order to identify the promoter/regulatory region of the peripheral and central part of the SA catabolic pathway, PCR - based on specifically designed primers - was used to amplify the intergenic region of various lengths preceding the two operons.

The fragments covering various parts of the regulatory regions were cloned into a promoter probe vector (pBBR1), upstream of the promoterless GFP reporter gene. The constructs were maintained in *Escherichia coli* and then introduced into *N. resinovorum* SA1 by electroporation. The GFP levels in these *N. resinovorum* SA1 strains grown in various culture conditions were followed with fluorimeter.

RNA samples used for RNA-Seq analyses were isolated from cultures grown on glucose or sulfanilic acid. Samples were taken from the cultures both at exponential and stationary-phases.

The cDNA libraries were produced from the isolated, ribosomal RNA depleted RNA samples and were sequenced on Illumina MiSeq platform.

Raw sequencing reads were trimmed and mapped to the protein coding regions of the *N. resinovorum* SA1 reference genome with the CLC Genomic Workbench 7 program.

Differential gene expression analysis was carried out with the EdgeR software package.

A selected representative set of the observed differential gene expressions were validated by means of RT-qPCR.

RESULTS

Novosphingobium resinovorum SA1 is a special strain because it can use sulfanilic acid as a sole carbon, nitrogen, sulfur and energy source without the assistance of any other organisms. I determined the complete genome sequence of the strain and identified the genes coding for enzymes taking part in the degradation process. I have resolved the phylogeny of these genes and studied the cellular processes controlling their expression. Moreover, applying RNA-Seq analyses I aimed to give a complex picture about the global effect of SA on the gene expression.

- I. Applying second generation sequencing technologies, the complete genome sequence of the *N. resinovorum* SA1 strain was determined. The bacterium possesses a complex genome composition. The genome consists of one chromosome and four additional plasmids (pSA1-4). Three of these plasmids are larger than 100 kbp and are considered as megaplasmids.
- II. The genes of the catabolic pathways of sulfanilic acid (SA), benzoate (B) and hydroxybenzoate (HB) degradation were identified on two large plasmids. The genes of the SA catabolic pathways forms two separate gene clusters on the pSA3 plasmid, while the genes of the B and HB pathway are located as separated operons on the pSA1 plasmid.
- III. The genes of the SA degradation pathway are partitioned into two gene clusters. The small one contains the peripheral part (SadAB), while the larger one harbors the central part of the pathway (ScaABCEF).
- IV. High degree of similarities were observed in the genomic organization and protein sequences between members of the SA catabolic pathway identified in *N. resinovorum* SA1 and other bacteria.
- V. According to the TBLASTN search in the NCBI databases, the SadA and SadB proteins of the *N. resinovorum* SA1 strain show the highest similarity to their homologs in the members of *Bradyrhizobiceae* family.
- VI. Another TBLASTN search using SadA and SadB proteins as query sequences in the sequence databases of *Comamonadaceae*, *Bradyrhizobiceae* and *Sphingomonadaceae* families showed that proteins similar to the key peripheral enzymes of the SA degradation are commonly present in the *Comamonadaceae* and *Bradyrhizobiceae*

strains and rarely occurring in the *Sphingomonadaceae* family which the *N. resinovorum* SA1 strain belongs to.

- VII. The components of the central part of the SA degradation pathway are less common among bacteria. They were identified exclusively in those bacteria, which were capable of sulfocatechol (SC) degradation.
- VIII. The correspondence analyses performed on the *N. resinovorum* SA1 genes showed that the codon preferences of the genes involved in SA degradation substantially differ from that of the other – e.g. housekeeping - genes. This is especially true for the genes of the peripheral pathway.
- IX. Based on the codon preference analyses, gene organization/genomic context, I concluded that emerge of the SA degradation pathway is a result of a recent horizontal gene transfer event.
- X. By means of modified promoter probe vectors (pPROBE), it was possible to map the genomic regulatory regions responsible for the expression of the peripheral and central part of the SA degradation pathway. My results show that the transcriptional regulations of peripheral and central part of the pathway substantially differ. The peripheral genes have high basal expressions, which further increase during starvation, the genes of the central part have basal expression levels which are specifically SA inducible.
- XI. The carbon catabolite repression influences the regulation of the central part of the pathway by blocking the SA induced expression in the presence of an easily consumable carbon source such as glucose.
- XII. RNA-Seq analyses revealed that growth phase (exponential vs. stationery) had larger impact on the global transcriptional landscape than the alteration of the carbon sources (Glc vs. SA).
- XIII. Gene set overrepresentation (GSOA) and gene set enrichment analyses (GSEA) proved that during the transition from the exponential to stationary growth phase, the differentially expressed genes mostly belonged to similar functional groups independently of the carbon sources used. In these cases, only downregulation could be observed which reflected the starvation effect on the cell. Notably, similar

starvation-related gene expression changes were observed when comparing exponential phase cells grown on different carbon sources.

- XIV. The transcript profiles of the genes taking part in the SA degradation supported the conclusions of the promoter mapping experiments. Genes of the peripheral part had high basal expressions, which further increased as a result of starvation. In contrast, the regulation of the genes of the central part was clearly SA dependent. Interestingly, the gene expression level of the common part of the pathway did not respond to the presence of SA or to starvation.
- XV. I could identify the elements of the electron transport chain linked to the peripheral pathway which were composed of a glutathione reductase (GR) and a plant type ferredoxin (Fer2) coding gene.
- XVI. Based on the transcriptome analysis, I identified a gene which encodes a transporter belonging to the MFS superfamily that is likely responsible for the SA uptake.
- XVII. Multiple genes were predicted to provide protection against sulfite released during the SA degradation. Three of them are sulfite transporters participating in sulfite transport into the periplasmic space and two of them are sulfite oxidizing enzymes which catalyze the oxidation of sulfite to sulfate in the periplasmic space.
- XVIII. The expression levels of the most important genes identified with RNA-Seq analyses were further confirmed by means of RT-qPCR technique.

PUBLICATIONS

Articles, the thesis based on:

1. **Botond Hegedüs**, Péter B. Kós, Gábor Bende, Naila Boundedjoun, Gergely Maróti, Krisztián Laczi, Márk Szuhaj, Katalin Perei, Gábor Rákhely: Starvation- and xenobiotic-related transcriptomic responses of the sulfanilic acid-degrading bacterium, *Novosphingobium resinovorum* SA1, Applied Microbiology and Biotechnology 102: pp. 305-318. (2018), **IF: 3,420**

2. **Botond Hegedüs**, Péter B. Kós, Balázs Bálint, Gergely Maróti, Han Ming Gan, Katalin Perei, Gábor Rákhely: Complete genome sequence of *Novosphingobium resinovorum* SA1, a versatile xenobiotic-degrading bacterium capable of utilizing sulfanilic acid, Journal of Biotechnology 241: pp. 76-80. (2017), **IF: 2,599**

Strongly related publications:

3. **Hegedüs Botond**, Kós Péter, Bende Gábor, Laczi Krisztián, Perei Katalin, Rákhely Gábor: Characterization of the biodegradation processes in *Novosphingobium subarcticum* SA1 grown on sulfonated aromatic compound in batch fermenter, Acta Microbiologica et Immunologica Hungarica 64:(suppl. 1) pp. 33-34. (2017)

4. Gábor Rákhely, **Botond Hegedüs**, Gábor Bende, Katalin Perei: Omics of sulfanilic acid biodegradation in *Novosphingomonas subarcticum*, New Biotechnology, 33: S129-S130 (2016)

5. Gábor Rákhely, Attila Bodor, **Botond Hegedüs**, Ágnes Kis, Krisztián Laczi, Gergely Maróti, Katalin Perei: How to eat toxic compounds? Metabolic insights into microbial degradation of xenobiotics, MTA SZBK Straub-napok 2016. május 25-26. (2016)

6. **Botond Hegedüs**, Zsolt Péntes, Gábor Bende, Katalin Perei, Gábor Rákhely: Replication and Partition System of the *Novosphingobium* Plasmids, Book of Abstracts of the 6th European Bioremediation Conference (ebc-vi 2015), 2015. p. 181. 1 p.

7. Gábor Rákhely, **Botond Hegedüs**, Ágnes Kis, Krisztián Laczi, Gábor Bende, Attila Bodor, Katalin Perei: Metabolic insight into biodegradation of sulfonated aromatic and other industrial hydrocarbon wastes, Book of Abstracts of the 6th European Bioremediation Conference (ebc-vi 2015), 2015. p. 256. 1 p.

8. Krisztián Laczi, Ágnes Kis, Balázs Horváth, Gergely Maróti, **Botond Hegedüs**, Katalin Perei, Gábor Rákhely: Metabolic responses of *Rhodococcus erythropolis* PR4 grown on diesel oil and various hydrocarbons, Applied Microbiology and Biotechnology 99:(22) pp. 9745-9759. (2015), **IF: 3,337**

9. **Botond Hegedüs**, Mónika Magony, Krisztián Laczi, Katalin Perei, Gábor Rákhely, András Tóth: Metabolic and protein-protein interactions involved in the sulfanilic acid degradation in *Novosphingobium subarcticum* SA1, Acta Microbiologica et Immunologica Hungarica 60:(Suppl.) p. 24. (2013)

10. **Botond Hegedüs**: The molecular mechanism of sulfanilic acid biodegradation, Acta Biologica Szegediensis 57:(1) pp. 86-87. (2013)

11. **Botond Hegedüs**, Katalin Perei, Mónika Magony, Krisztián Laczi, András Tóth, Kornél L. Kovács, Gábor Rákhely: Metabolic and protein-protein interactions of sulfanilic acid catabolism in *Novosphingobium subarcticum* SA1, Environmental Engineering and Management Journal 11:(3/Suppl.) p. S20. (2012)

12. Gábor Rákhely, **Botond Hegedüs**, Mónika Magony, Krisztián Laczi, András Tóth, Gergely Maróti, F K Medzihradzsky, Kornél L. Kovács, Katalin Perei: Metabolism of sulfonated aromatic compounds in *Novosphingobium subarcticum* SA1 strain, Environmental Engineering and Management Journal 11:(3/Suppl.) p. S5. (2012)