

Szent István University

Methods of enzyme analyses: from gene dosage change to structure determination

Or the determination of the thioesterase II gene role in Bacitracin synthesis and analysis of plant cell wall degrading enzymes

PhD thesis abstract

Lóránd Szabó

2006.

PhD school:

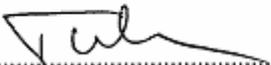
name: Biological Sciences PhD school

branch: biological sciences

school leader: Prof. Dr. Zoltán Tuba, DSc.
Szent István University, Faculty of Agricultural and Environmental Sciences
Department of Botany and Plant Physiology

Project supervisors: Prof. Dr. Tibor Sík, DSc
professor emeritus
Eötvös Lóránd University, Faculty of Science, Institute of Biology
Department of Genetics

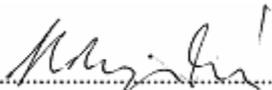
Dr. András Holczinger
Eötvös Lóránd University, Faculty of Science, Institute of Biology
Department of Genetics
Budapest University of Technology and Economics
Department of Agricultural Chemical Technology – Fermentation Pilot Plant
Laboratory


.....

School leader


.....

Project supervisor


.....

Project supervisor

Preface

The thesis consists of two separate parts due to a nearly two years long work in abroad. The first part describes the work done at the Biotechnology and Molecular Genetic Department of Szent István University, Hungary, which was the cloning, and characterization of an essential enzyme of the Bacitracin biosynthesis. The second part is the summary of the work carried out with plant cell wall degrading enzymes at the Department of Biological and Nutritional Sciences of The University of Newcastle upon Tyne, in the UK.

The experimental work was accomplished by the end of 2000 and the first version of this thesis was prepared in 2001. After a long pause it was completed by the end of 2006. Due to this, the literature review and the discussion contain references only up to 2001.

1. Determining the role of the thioesterase gene in bacitracin biosynthesis

1.1. Introduction and research objectives

The fast development of molecular genetic methods changed the way of the new active pharmaceutical ingredient identification. Instead of the earlier used try-and-error method a more rational process was developed. First the molecular mechanism of the illness is identified (pathogenic microorganism, not proper expression or activity of an enzyme or regulator, etc.). In the next step high throughput screening methods are applied to identify those molecules, which do have effect on the identified mechanism itself. In this modern drug research method the number of testable complex chemical structures became the tight cross section of the new drug screening process.

Many complex chemical structures do exist where the chemical synthesis is uneasy or impossible but can be produced biologically, by fermentation in huge quantity. Polyketides and polypeptides belong to this class of compounds. To be able to ferment these complex chemical structures according to our needs the biosynthetic routes of the production must be know in great details.

The bacitracin peptide antibiotic is produced not on ribosomes, but by a three-domain peptide synthetase via nonribosomal peptide synthesis. The main steps of the nonribosomal peptid synthesis catalyzed by the peptid synthetases are already clarified, although the determination and characterization of the other essential enzymes required for the process are still under investigation.

The main aim of this work was to localize and characterize an essential enzyme in the nonribosomal peptide synthesis located around the coding region of the bacitracin synthetase.

1.2. Literature review

For a long time the bacitracin peptid antibiotic was used as an animal feed additive, but after the withdrawal of antibiotics from the animal feeds its human application got wider.

The bacitracin peptid antibiotic complex is synthesized by cells of *Bacillus licheniformis* after the onset of the stationary phase. The chemically well-defined bacitracin peptid contains 12 amino acids and synthesized by the bacitracin multienzyme complex in a more than 50 step process (Laland and Zimmer, 1973).

1.2.1. Peptid biosynthesis by multienzyme complexes

During multienzyme complexes catalyzed peptide biosynthesis each amino acids of the target peptide are introduced by a 1000 amino acid sized module. These modules can be further divided into domains which catalyze the substeps and separated from each other by flexible linker region. These domains are: adenylation domain (A), peptid carrier domain (PCP or T), modification domain (M) and the condensation domain (C). The readily made peptid is separated from the

multienzyme complex by the thioesterase domain (Fig. 1.2.1.1.). (Laland and Zimmer, 1973; Aharonowitz et al. 1993; Stachelhaus and Marahiel, 1995a/b; Stein et al. 1996; Marahiel et al. 1997; van Wageningen et al. 1998; Cane and Walsh, 1999).

The adenylation domain is responsible for the specific recognition and activation of the amino acids. The so far identified adenylation domains were classified based on amino acid homology and properties of the substrate binding pocket. (Stachelhaus et al. 1999; Challis et al. 2000) Using this classification the amino acid specificity of the newly identified adenylation domains can be forecasted.

The activated amino acids are covalently linked through a thioester bond to the 4' phosphopantetheine arm of the PCP domain. The function of the PCP domain is similar to the ACP domains of polyketide and fatty acid synthases, since it carries amino acids or peptide chains between catalytic centers. The 4' phosphopantetheine arms are covalently linked to a serine residue of the PCP domain by 4'PP transferases in a process which releases CoASH and uses ATP (Lambalot et al. 1996).

The modification domain is not part of the minimal module, so it is present only in those modules where the amino acid is modified before introduced into the peptide chain. Mostly epimerisation happens, which converts the amino acid from the L- into the D-form. Heterocycle ring can be formed between cysteine and serine or threonine side chains by the heterocyclase domain (Konz et al. 1997). In eukariotes the methylation domains carry out the methylation of the α -carbon amino groups (Zocher et al. 1986).

The condensation domain forms the peptid bond. The first module of the peptid synthetases does not contain condensation domain, these modules are called initialization modules. The condensation domain containing elongation modules are not able to initiate peptide bond formation as long as the acceptor site of the condensation domain is not filled with an amino acid activated and linked by the upstream module (Stachelhaus et al. 1998).

The 250 amino acid sized thioesterase domains detaches the mature peptide chain from the last module PCP domain 4' phosphopantetheine arm (Schneider és Marahiel, 1998; Shaw-Reid et al. 1999).

1.2.2. The bacitracin synthetase

The 12 amino acid containing branched, peptide antibiotic bacitracin is synthesized by *Bacillus licheniformis* cells after the onset of the stationer phase. Bacitracin A is the major component (Fig 1.2.2.1.) (Ikai et al. 1994; Morris, 1994; Siegel et al. 1994; Epperson and Ming, 2000). The minor components differ from the major component bacitracin A only at positions X and Y and at the first amino acid of the thiasoline ring (Fig. 1.2.2.1.). The amino thiasoline ring containing forms are biologically actives (Fig. 1.2.2.1. A and B) but the keto-thiazolin ring containing forms are biologically inactive (Fig. 1.2.2.1. C and D).

The bacitracin dodecapeptides are synthesized by a three component multienzyme complex called bacitracin synthetase. Using bacitracin non-producer Tn917PF1 transposone mutants the partial physical map of the bacitracin synthetase coding region was determined (Prágai et al. 1994b) and the promoter of multienzyme coding region was isolated and the transcription initiation site was determined using primer extension (Prágai, unpublished).

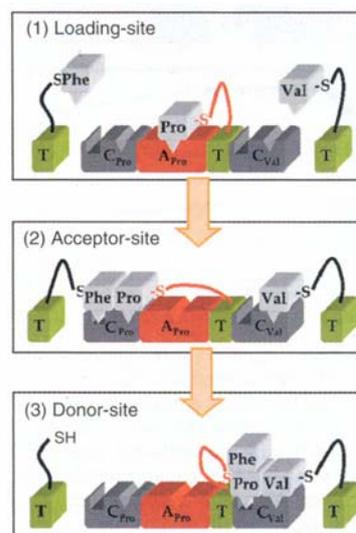


Fig. 1.2.1.1. The process of nonribosomal peptide synthesis (Belshaw et al. 1999)

Bacitracin	X	Y	1 st position	R
A	L-Ile	L-Ile	L-Ile	A
B ₁	L-Ile	L-Ile	L-Val	B
B ₂	L-Val	L-Ile	L-Ile	A
B ₃	L-Ile	L-Val	L-Ile	A
D ₁	L-Val	L-Ile	L-Val	B
D ₂	L-Ile	L-Val	L-Val	B
D ₃	L-Val	L-Val	L-Ile	A
E	L-Val	L-Val	L-Val	B
F	L-Ile	L-Ile	L-Ile	C
H ₁	L-Ile	L-Ile	L-Val	D
H ₂	L-Val	L-Ile	L-Ile	C
H ₃	L-Ile	L-Val	L-Ile	C
I ₁	L-Val	L-Ile	L-Val	D
I ₂	L-Ile	L-Val	L-Val	D
I ₃	L-Val	L-Val	L-Ile	C

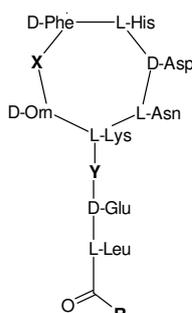
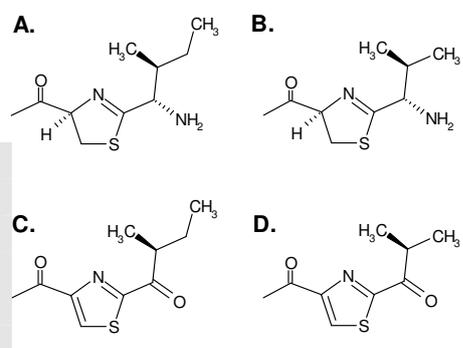


Fig. 1.2.2.1. The bacitracin peptid antibiotic complex



The bacitracin synthetase coding 45 kb long DNA fragment was cloned and sequenced (**Konz et al. 1997**). (**Fig. 1.2.2.2.**) Downstream of the peptide synthetase coding regions (*bacA*, *bacB*, *bacC*) ORFs coding a two component regulator system were identified (*bacR*, *bacS*). (**Podlasek et al. 1995**). Downstream to the two component regulator an ABC transporter coding region (*bcrA*, *bcrB*, *bcrC*) was identified which was shown to be responsible for the bacitracin resistance. Knock out of the sensor component of the two component regulator resulted increased bacitracin sensitivity of the cells (**Neumuller et al. 2001**).

Amplification of the bacitracin resistance coding region resulted not only increased bacitracin resistance but also increased bacitracin production in some cases (**Podlasek et al. 1995**)

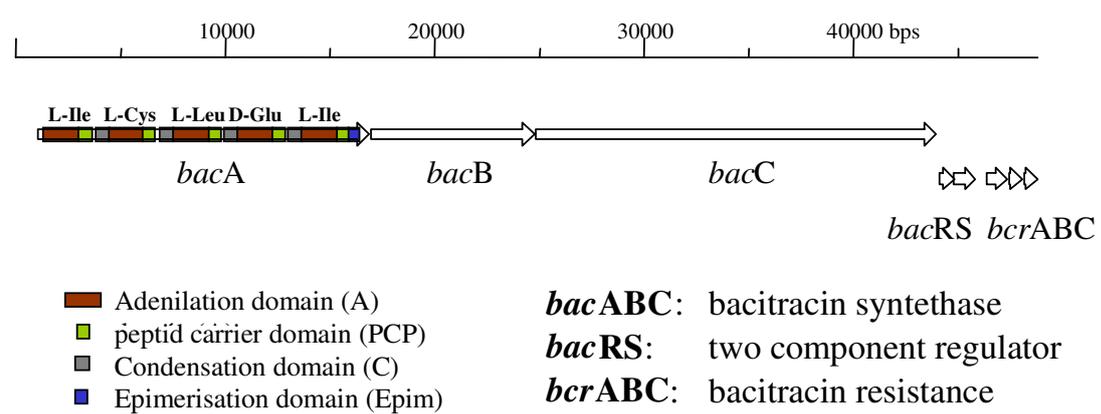


Fig. 1.2.2.2. Structure of the bacitracin synthetase coding region (AF007865)

1.3. Materials and methods

In the experiments commercially available materials were used (SIGMA, Reanal). The used *E. coli* strains were *E. coli* DH5 α and *E. coli* XL1-Blue. The *Bacillus licheniformis* strains (*B. licheniformis* 19, *B. licheniformis* 19F4) were identified earlier in our laboratory and used for commercial production of bacitracin at Phylaxia. The basic plasmids are either earlier described

(PQF50) or commercially available (pBlueScript II KS). The used constructs were generated from the basic plasmids using standard molecular techniques. The *Bacillus licheniformis* protoplasts were transformed as described by Prágai et al, 1994/a. The β -galactosidase activity of the *lacZ* fusion constructs were determined using the method of Nicolson and Setlow (1990). Sequence analysis was done by using the GCG Wisconsin Package, Version 8 software.

1.4. Results

The DNA fragment coding the promoter of the bacitracin synthetase cloned from a bacitracin non-producer Tn917PF1 transposon inactivated mutant was already available. (Prágai, unpublished) (Fig. 1.4.1.). The so far unanalyzed upstream, *Bam*HI-*Sma*I region of the fragment was subcloned into the pBlueScript II KS vector and sequenced. On the 705 bp long DNA fragment an ORF was identified. The

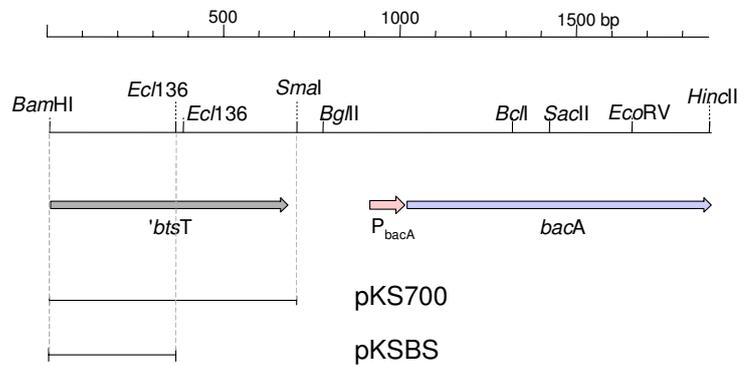


Fig. 1.4.1. The used DNA fragment and the origin of the subclones

ORF coded protein showed high homology to eukaryotic thioesterase II enzymes and other prokaryotic proteins which were always localized close to peptide synthetase coding regions. Due to the high homology of the coded protein to thioesterases the ORF was named *orfT*. Based on the sequence alignments the cloned fragment did not contain the start of the ORF.

To identify the function of *orfT* a construct was prepared which was able to inactivate the intact, chromosome located *orfT* gene upon integration (Fig. 1.4.2. case A). Protoplasts of *Bacillus licheniformis* cells were transformed with the inactivating construct containing an inside fragment of the *orfT* gene and a kanamycine resistance cassette. Bacitracin production of the transformant colonies was measured biologically using *Micrococcus flavus* as a test microorganism. Bacitracin production of the

transformants was only 10 % of the wild type strain. This result showed that the *orfT* is important in bacitracin production, and it was renamed to *btsT*.

Since the inactivation of the *btsT* ORF (which means the loss of BtsT enzyme) resulted severe decrease of bacitracin production the following question emerged: What happens if the concentration of the BtsT enzyme is increased? One of the easiest ways of increasing the concentration of an enzyme is when the enzyme coding gene is introduced into the cell on a high copy number plasmid. To achieve this: a high copy number plasmid, the whole *btsT* ORF and a promoter driving the *btsT* ORF was required.

To clone the whole *btsT* ORF genome walking technique was applied. The newly cloned region was sequenced. Using the alignment of other homologous thioesterase proteins the position of the ATG and RBS (ribosome binding site) was determined (Fig. 1.4.3.)

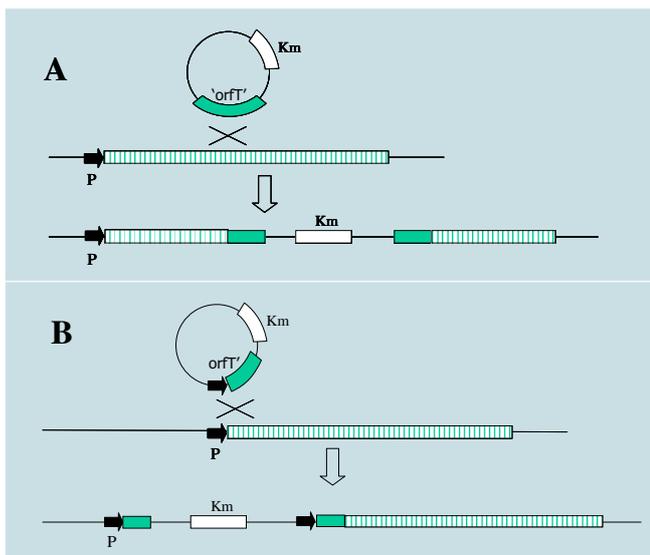


Fig. 1.4.2. Cases of single plasmid integration

- A: the gene is inactivated if the integrative construct contains the inside part of the gene
 B: if the construct contains either end of the gene an intact gene will be generated

RBS

1 CGGACGGCTG ACGTATCCGT CGATTAACGG GTTATTCACC AAAAAATATC TACAACGAAA

MetLysLeu PheCysLeu ProTyrAla GlyGlySerGlu SerAlaPhe·

61 GGGAAATGATA GAAATGAAAT TATTTTGCCT GCCTTACGCC GGCGGATCCG AATCCGCATT

·TyrSerTrp LysGlyHisMet GlnProAsp IleGluIle CysProIleGln LeuLysGly·

121 TTATTCCTGG AAAGGCCATA TGCAGCCCGA CATTGAGATT TGTCCGATTG AGCTGAAGGG

·ArgGlyArg ArgPheAsnGlu ProCysTyr GluSerLeu GluGluAlaVal GlnAspIle·

181 AAGGGGCCGG CGTTTCAACG AGCCTTGTTA CGAAAGCCTT GAAGAAGCAG TTCAAGACAT

·PheGluGln ValGlnAlaGlu ArgLysGly AspAspTyr ProLeuPheGly HisSerMet·

241 TTTTGAGCAG GTTCAAGCTG AACGAAAAGG TGACGACTAC CCTCTTTTTCG GGCACAGCAT

·GlySerLeu LeuAlaTyrGlu LeuTyrTyr GlnMetSer GlyAlaGlyAla GluLysPro·

301 GGGAAAGCCTT TTGGCATATG AACTTTACTA TCAAATGAGC GGGGCGGGAG CTGAAAAACC

·ValHisIle PhePheSerGly TyrLysAla ProAsnArg IleArgLysThr GluLysLeu·

361 GGTTCACATT TTTTCTCGG GCTATAAAGC GCCAAACAGG ATCAGAAAGA CAGAAAAACT

·HisThrLeu ProAsnProIle PheLysLys LysIleVal GluLeuGlyGly ThrProGlu·

421 GCATACCTTG CCCAATCCTA TTTTAAAGAA AAAAATTGTC GAGCTCGGGG GAACGCCTGA

·GluLeuIle AsnHisGluGlu LeuPheGlu LeuPheIle ProIleLeuLys SerAspPhe·

481 GGAGCTCATC AATCATGAAG AGCTATTTGA ATTGTTTATC CCCATTCTCA AAAGCGACTT

·LysMetVal GluAsnTyrIle TyrGlnGlu ArgAsnSer LysIleAspCys AspIleThr·

541 TAAAATGGTA GAAAACATA TCTATCAAGA AAGAAACAGC AAAATAGATT GCGACATTAC

·ValLeuAsn GlyLysGluAsp AlaMetSer LysGluHis ValSerAspTrp LysHisHis·

601 CGTTCTCAAC GGAAAAGAAG ACGCCATGAG CAAGGAACAT GTATCCGATT GAAACATCA

·ThrSerGly HisPheThrAla TyrTyrPhe GluGlyAsn HisPhePheLeu HisHisHis·

661 TACTTCAGGA CACTTTACAG CCTATTACTT TGAGGGGAAT CATTCTTTT TGCACCATCA

·ValGluLys IleThrGluIle IleAsnHis SerLeuThr AlaSerArgThr Phe***

721 CGTTGAAAAG ATCACCGAAA TCATCAATCA TTCACTGACA GCCAGCCGGA CGTTTTAACC

781 TGCGATTTCG GCGAGATTCA AGCCCGGG

Fig. 1.4.3. Sequence of the *btsT* gene (AF050160)

Using Campbell type recombination it was proved that the newly cloned DNA fragment contained not only the whole *btsT* ORF, but also a promoter upstream by using a construct containing the upstream end of the *btsT* gene. This construct would have disrupted bacitracin synthesis if the cloned fragment would not contain the upstream part and the promoter of the *btsT* ORF.

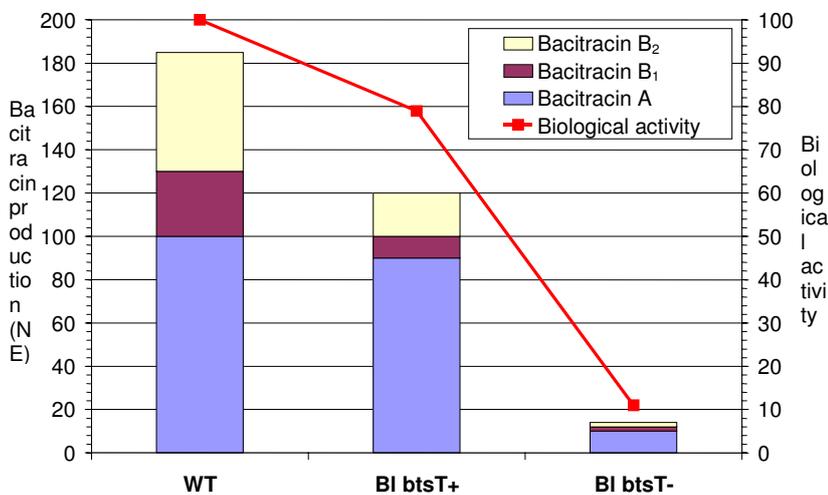


Fig. 1.4.4. Bacitracin production determined by bioassay and HPLC

To test the *btsT* gene amplification effect on bacitracin synthesis, the *btsT* gene with its own promoter was transferred into the middle copy, *Bacillus* vector pUB110. The generated construct was used to transform *Bacillus licheniformis* protoplasts. Bacitracin production of the plasmid construct containing strains (BIbtsT+) was compared with the bacitracin production of the above described *btsT* inactivated strain (BIbtsT-) and the wild type strain (WT) using

bioassay and HPLC measurement of the bacitracin complex (**Fig. 1.4.4.**). Results showed that the biological activity of the produced bacitracin was nearly 20 % lower of the BlbtsT+ strain compared to the wild type strain. Based on the HPLC measurement of the three most significant bacitracin factors (A, B₁ and B₂) it was clearly seen that the reduction in biological activity of BlbtsT+ strain produced bacitracin is due to the significant decrease of the B₁ and B₂ factors while the biologically most active bacitracin A factor was only slightly decreased.

To test the transcription of the *btsT* gene and the subunits coding region (*bacA*) such kind of endogen β -galactosidase enzyme non-producing *Bacillus licheniformis* 19F strains were prepared where a promoterless β -galactosidase reporter gene was introduced into the *btsT* and *bacA* DNA fragments, respectively (**Fig. 1.4.5.**)

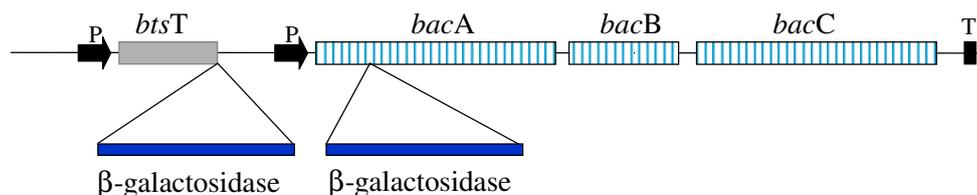


Fig. 1.4.5. Insertion points of the β -galactosidase reporter gene into the *btsT* (thioesterase) and *bacA* (peptide synthetase first subunit) region P: promoter; T: terminator; β -galactosidase

Reporter gene containing mutants and wild type *Bacillus licheniformis* 19 cells were cultured and sampled in hour intervals. From the samples the growing was measured and bacitracin production was determined using bioassay. The β -galactosidase reporter gene activity was quantified using ONPG (**Fig. 1.4.6.**)

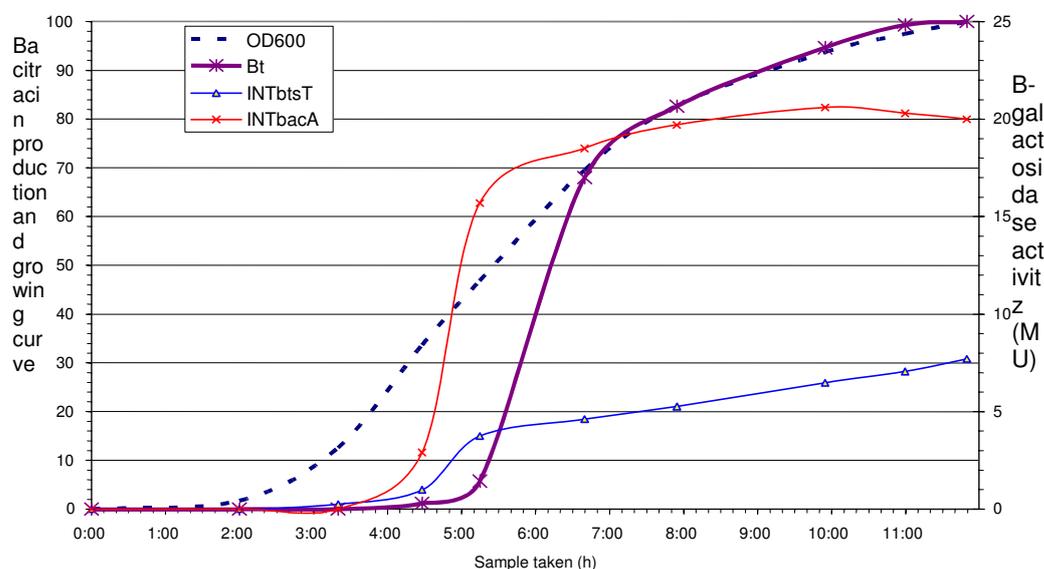


Fig. 1.4.6. Transcription of the *btsT* and *bacA* genes based on β -galactosidase reporter gene assays
INTbtsT: reporter gene in *btsT*; INTbacA: reporter gene in *bacA*; OD₆₀₀: growing curve (% of the maximum); Bt: bacitracin production (percentage of the maximum)

1.5. Discussion

The bacitracin peptide contains a chemically non-synthesizeable thiasoline ring (**Fig 1.2.2.1**). Due to this the bacitracin synthetase become interesting and was sequenced by others soon (**Koncz et al. 1997**).

Before the publication of the sequence of the bacitracin synthetase coding region five bacitracin non-producer mutants were isolated in our lab. Using these transposon inactivated non-producer strains a partial physical map of the bacitracin synthetase coding DNA region was developed (**Prágai et al. 1994/b**). From one of the mutants a DNA fragment was isolated which was shown to contain the promoter and one kilobase of the coding sequence of the first bacitracin synthetase subunit. The promoter region was localized, sequenced and characterized using primer extension (**Prágai, unpublished**). Since the sequence of the bacitracin synthetase coding DNA was published and the cloning work done till that time lost its novelty, our attention was focused on the promoter upstream region which was not analyzed so far (**Fig. 1.4.1.**)

The promoter upstream region was cloned and sequenced. On the fragment an ORF was identified which coded a protein showing high homology with eukaryotic thioesterase II proteins and enzymes that were very often coded around peptid synthetases coding DNA fragments. Based on the homology with thioesterases the ORF was named *orfT*.

Eukaryotic thioesterase II are expressed specifically only in certain tissues like lactating mammary gland of nonruminant animals and uropygial gland of waterfowl. These thioesterase II enzymes produce shorter chain fatty acids (C8-C12) instead of the standard C16-C18 size. Under standard conditions the C16-C18 size fatty acids are separated from the synthetase by the thioesterase I module. The thioesterase I and II proteins has amino acid similarity only at the active site (**Libertini and Smith, 1978; Wakil et al. 1989; Pazirandeh et al. 1989; Pazirandeh et al. 1991**).

The rat thioesterase II enzyme was cloned and biochemically analyzed and the catalytic amino acids were identified using site specific mutagenesis (**Tai et al. 1993**). The TEII catalytic mechanism is analog to the serine proteases catalyzed reaction.

The function of the other homolog group, the around peptide synthetase coding region located TEII was not determined at that time.

Inactivation of the *orfT* significantly reduced the bacitracin biosynthesis. The promoter of the bacitracin synthetase coding fragment was localized right upstream of the gene, downstream from the thioesterase coding region. *In silico* a terminator like hairpin structure could be identified between the thioesterase and the synthetase coding genes. Based on these findings the decrease of bacitracin production is not due to the separation of the promoter and synthetase coding region by the integration, but due to the essentiality of the BtsT protein in the bacitracin biosynthesis. Similar results were found when the TEII coding gene of the surfactin synthetase was inactivated (**Schneider és Marahiel, 1998**). Due to the similarity to the TEII proteins and essentiality in the bacitracin biosynthesis the *orfT* gene was renamed to *btsT*.

The effect of the *btsT* gene dosage on bacitracin synthesis was tested by comparing the bacitracin production of the wild type (WT), the inactivated *btsT* (BlbtsT-) and amplified *btsT* (BlBtsT+) gene containing strains (**Fig 1.4.4.**).

Results showed that the biological activity of the produced bacitracin reduced in both cases. The BlbtsT+ strain produced bacitracin showed 80 % biological activity compared to the wild type, while the BlbtsT-, the *btsT* inactivated strain, produced bacitracin of 10 % biological activity compared to the wild type. The HPLC measurement of the bacitracin complex showed that while all bacitracin components of the BlbtsT- strain showed severe reduction, in the BlbtsT+ strain only the production of the minor B₁ and B₂ forms were reduced.

The bacitracin A and B₁ and B₂ forms differ only in one amino acid. The B₁ form at the first position, the B₂ form at the 8th position contains valine instead of the bacitracinA form present isoleucine. The valine and isoleucine differs only in the size of the amino acid specific side chain, isoleucine has one CH₂ group longer side chain.

Adenylation domains of the bacitracin synthetase responsible for introducing the first and the 8th amino acids has similar substrate specificity determining amino acids in the substrate binding cleft as the FenB and LicC adenylation domains. FenB and LicC were already shown to be able to introduce valine instead of isoleucine in *in vitro* experiments (**Doekel and Marahiel, 2000**). Based on this the appearance of the valine instead of the isoleucine and so the formation of the bacitracin B₁ and B₂ form is due to the “low” substrate specificity of the first and 8th adenylation domain of the bacitracin synthetase.

It was noticed that even if the substrate activation of the isolated adenylation domain is not specific *in vitro*, the *in vivo* produced final product of the peptide synthetases does not contain wild range of amino acids. This is explained by the fact that the substrate specificity of the condensation domain is high, so condensation domains are not able to covalently link any amino acids to the growing peptide chain. So the activated amino acids which are not proper substrate of the condensation domains stuck the peptide synthesis as long as not removed and replaced by the proper amino acid.

Since the increase of the copy number of the *btsT* ORF and its protein product reduced mostly the amount of the B₁ and B₂ bacitracin form without significantly changing the A form, it is obvious that the BtsT enzyme is responsible for the removal of the not properly activated amino acids from the 4' phosphopantetheine arm.

The effect of the absence or increased presence of the BtsT enzyme is explained as follow:

- if the BtsT enzyme is missing (BIBtsT-): the peptide synthesis is stopped when the first amino acid is activated which is not acceptable substrate of the condensation domain. Although in the cell new peptid synthetases are expressed, eventually all of them are stuck, so the total produced bacitracin by the culture will be low.
- if the BtsT enzyme is overproduced (BlbtsT+) removal of the non-proper substrate of the condensation domains has increased chance, so mostly those amino acids are incorporated into the final product which can be processed fast by the condensation domain

Based on our work the thioesterase II enzymes coded in the surrounding of the prokaryotic peptide synthetases are responsible for the removal of the mistakenly adenylylated amino acids which can not be incorporated into the growing peptide chain by the condensation domain.

Although no structure information is available of the thioesterase II proteins, it is still possible to determine the overall fold of these proteins. First, based on the alignment of several thioesterase II proteins it is possible to determine which the most favored amino acids of each position are. Then based on the amino acid distribution the secondary structure of each position can be forecasted. By comparing the order of the secondary structure elements with known structures the overall fold can be determined. Based on SAM results (Sequence Alignment and Modeling Software System) (**Hughey and Krogh, 1996; Karplus et al. 1999; Krogh et al. 1994**) the fold of the BtsT protein belongs to the α - β fold group (**Fig. 1.5.1.**). The α - β proteins are further divided into subgroups based on the deviation from the general α - β fold (**Heikinhemino et al. 1999**). The BtsT protein belongs to the HAL (haloperoxidase) subgroup, but the β 1 and β 2 sheets are missing and the region between the β 6 and α D helix covering the substrate binding pocket are different. The amino acid similarity between the haloperoxidase and BtsT protein is neglectable except the active site.

The rat's TEII protein was investigated using site-specific mutagenesis (**Tai et al, 1993**). Based on these experiments the Ser101, His237 are the active site residues. The changing of the third component of the triad from Asp236 to Ala236 did not influence the TEII catalyzed reaction. From this it was concluded that acid component of the triad (differently from the proteases) is not required for the reaction.

Based on the forecasted fold of the BtsT protein, the conserved position of the α - β hydrolase fold catalytic amino acids and the position of the conserved amino acids of the thioesterase II group the catalytic amino acids of the BtsT protein were be determined. Ser75, which is located in the highly conserved nucleophylic elbow, is the nucleophyl, His208, located on the loop between the β 8 sheet and α F helix is the proton holding histidine, and Asp181 is the histidine stabilizing acid component located on the loop between β 7 and α E.

Alignment of the rat TEII protein and BtsT protein showed the earlier tested Asp236 amino acid of the rat thioesterase protein ((**Tai et al. 1993**)) is not part of the triad. This is the reason why its change to alanine did not abolish thioesterase activity of the rat thioesterase. Based on the catalytic triad determined for the BtsT protein the catalytic triad of the rat thioesterase II protein is consist of the Ser101, Asp185 and His237 amino acids.

Using endogen β -galactosidase non-producing *B. licheniformis* F strain we were able compare the transcription of the thioesterase gene and the peptide synthetase first subunit-coding region by integrating promoterless β -galactosidase reporter genes into the appropriate ORFs. The proliferating cultures were sampled and the cell-growing, the bacitracin production and β -galactosidase activity

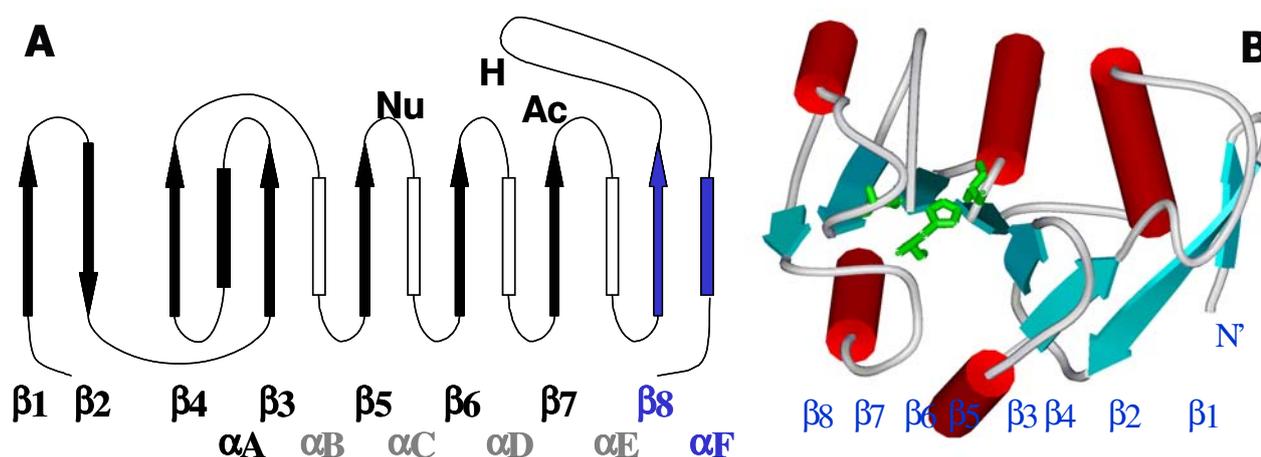


Fig. 1.5.1. α - β hydrolase fold. **A**: schematic drawing: black: minimal fold; black+blue: canonical fold; white: possible extensions. Arrows: β -sheet; boxes: α -helix; Nu: nucleophile (S/C); Ac: acid (E/D); H: histidine. **B**: 3D structure: (rotated by 180° compared to the schematic drawing) blue arrows: β -sheet; red cylinder: α helix; green: amino acids of the catalytic triad.

was measured. Based on the results (**Fig. 1.4.6.**) the transcription of synthetase coding region starts in the exponential phase and reaches the maximum in one hour. The expression of the thioesterase gene follows the same pattern, although the maximum of the expression level is only 25-30 % of the synthetase coding gene's transcription. The bacitracin production starts when the cells enter into the stationer phase, when the transcription from the tested genes is maximal.

In industrial application during the fermentation of the target molecules very often, molecules with more or less similar structure are also synthesized. The separation of these analogs from the target product could be very difficult during the downstream processing. These analogs are called "impurity". The removal of these impurities is very often require modification either in the fermentation or in the DSP (downstream processing) process making the production more expensive and less efficient. During the work we showed that over expression of the thioesterase II enzyme reduced the production of the bacitracin B₁ and B₂ impurities without significantly changing the target bacitracinA production, so the thioesterase II enzyme could be an important tool to reduce impurity during strain improvement projects.

2. Plant cell wall degrading enzymes

2.1. Introduction and research objectives

2.1.1. Role of the plant cell wall

The plant cell wall is not only the significant part of the produced biomass, and the only form of digestible fiber but also a renewable energy source. Only from cellulose (which is only 20-30 % of

the plant cell wall) 4×10^{10} tones are produced on yearly bases (Coughlan, 1985). Beside the cellulose the standard plant cell wall contains 50-60 % non-cellulose polysaccharides: pectin, hemicellulose and glycoprotein (Albersheim et al. 1994; Knox et al. 1990). Due to the bacterial and fungal plant cell wall degrading enzymes this multifunctional, complex and stable material is not accumulated, but are degraded into small, utilizable compounds. The plant cell wall degrading enzymes are not only for reintroducing carbon in the carbon cycle but also used by the food industry, garment trade and paper industry.

Since all application requires different properties, the enumeration of the enzymes and understanding their mechanism is required before the application specific changes can be made.

2.1.2 Structure of the plant cell wall

Plant cell wall is a complex structure which is composed from the cellulose micro-fibers and the matrix which connects the fibers and fills the space among (Fig. 2.1.2.1.). The matrix is a complex of several different compounds, which can be grouped according to the method of extraction from the plant cell wall. Pectins can be extracted with hot, diluted acids. The follow up alkali treatment extracts the hemicellulose. The remaining frame is the cellulose (Dey and Brinson, 1984). In some cases when the growing of the plant cell wall is finished lignin is built into the walls which strengthen it (Varner and Lin, 1989).

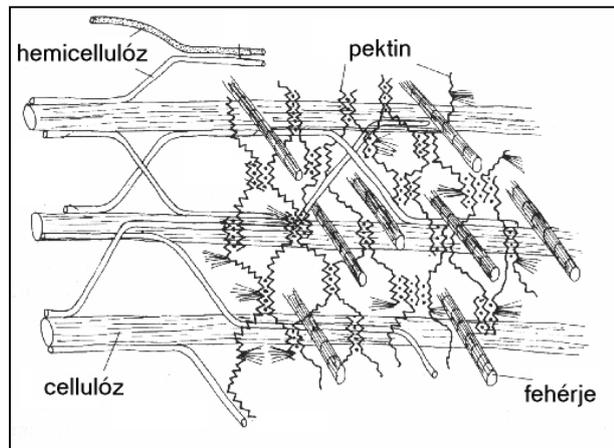


Fig. 2.1.2.1. Structure of the plant cell wall

2.1.2.1. Structure of the cellulose

Cellulose is a chemically simple homopolymer of β -1,4 linked glucose. Each glucose molecule is rotated by 180° compared to the previous. The two glucose containing, repeating unit is a disaccharide: the cellobiose. The glucose rings of the chain are positioned along a surface. In the micro-fibers the cellulose chains are arranged parallel and connected by inter and intra-molecular hydrogen bonds and van der Waals forces (Coughlan, 1985).

2.1.2.2. The matrix

In the Ferro-concrete structure like plant cell wall the Ferro part is composed up by the cellulose micro-fibers. The space between the fibers is filled by the amorphous matrix which is the complex mixture of hemicelluloses, pectins, glycoproteins and sometimes lignins (Albersheim et al. 1994).

Hemicellulose is the most significant non-celluloid component of the cell wall which bind to the cellulose by hydrogen bonds and keeps the micro-fibers in position till the end of the lignification. Hemicellulose is chemically much more complex than the cellulose and often contains branches and side chains. Due to these, its structure is not that ordered as cellulose. Hemicelluloses are divided into subcategories based on the main chain sugar: xylan, mannan, galactan, glucomannan, xyloglucan, callose, and arabinogalactan (Brett and Waldren, 1996).

Xylan is a homopolymer of β -1,4 linked xylose, which builds up 30 % of the plants' dry weight (Joseleau et al. 1992). Beside the xylose xylans often contains glucuronic acid, acetyl and arabinose side chains. (Dekker and Richards, 1976; Biely et al. 1985; Poutanen, 1988; Kormelink et al. 1993).

Mannan is a polymer of β -1,4 linked mannose. In the galactomannan galactose units are linked to the mannan frame by β -1,6 links. The mannan without side chains could be crystalline, and in case of some green algae, it can replace cellulose (Yui et al., 1997).

Galactomannans has a much open structure that can hold huge amount of water, which can be important at the seed germination process (**Brett and Waldren, 1990**).

Glucomannan is the polymer of β -1,4 linked glucose and mannose. In gymnosperma β -1,6 linked galactose could be present. This gluco-galacto-mannan is more water soluble than the glucomannan. At the C2 or C3 position of the mannose units acetyl groups can be present (**Hazlewood and Gilbert, 1998a/1998b**).

In case of **xyloglucan** the β -1,4 linked glucose side chain contain α -1,6 linked xylose units. To these xylose units very often fucose, galactose, arabinose side chains are linked by α -1,2 bonds (**Brett and Waldren, 1996**).

During the plant wound healing β -1,3 linked glucose is formed which is named **callose**. Callose has a helical structure so it can make up micro-fibers or can be gelatinized in the presence of water.

Arabinogalactan is built up from β -1,3 and β -1,6 linked galactose units.

Pektins are complex acidic polysaccharides built from galacturonic acid, rhamnose, arabinose and galactose. Rhamnogalacturonan, araban, galactan and arabinogalactan belongs to this group. Pectins build up their own texture and linked to the cellulose micro-fibers and hemicellulose by covalent and hydrogen bonds. Pectins contain smooth and hairy regions depending on the length of the side chains.

Plant cell wall also contains **proteins**, which are very often glycosylated. These proteins often contain hydroxyproline and called hydroxyproline rich glycoproteins (HRGPs). The plant cell wall contains peroxidase, invertase, cellulase, phosphatase, pectinase, pectin methyltransferase, malate dehydrogenase, exoglycosidases, endoglycosidases and endotransglycosidases. These enzymes are required for the maintenance of the plant cell wall structure (**Brett and Waldren, 1990**).

Lignins are hydrophobic, unordered polymer of different, covalently linked alcohols, which are built into the cell wall after the cells stopped growing. The polymerization is not enzyme catalyzed, so the filling in of the space continues as long as all the material is used up and the space is completely filled. After lignin built up growing of the cell is not possible. Lignin also protects the cell from intruders. Lignified cells are always dead and protect the inside cells (**Brett and Waldren, 1990**).

2.1.3. General structure of the glycoside hydrolases

Glycoside hydrolases has modular structure (**Fig. 2.1.3.1.**) since these enzymes contains not only the catalytic domain but also several other modules with different function (**Tomme et al. 1995; Warren, 1996**).

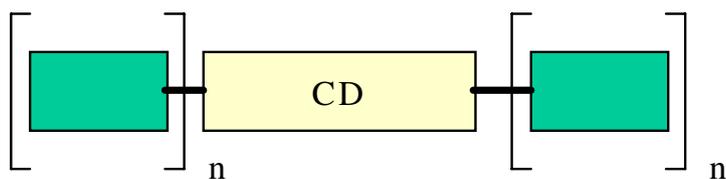


Fig. 2.1.3.1. General structure of glycoside hydrolases. CD: catalytic domain; black lines: linker regions; green boxes: other modules (usually carbohydrate binding modules)

Warren, 1996).

The modules linked to each other by flexible linker regions. Since the linker regions are easily available for the proteases, the modules can be separated by limited proteolysis (**Gilkes et al. 1988; Tomme et al. 1988**). In most cases the catalytic domain linked modules are

carbohydrate binding modules (CBM) but can have esterase, thermo-stabilizing, bacterial cell wall binding, linker or dockerin function (**Henrissat and Coutinho, 2000**).

2.1.3.1. Catalytic domain of the glycoside hydrolases

Glycoside bond hydrolyzing catalytic domains were grouped into 85 families based on amino acid sequence homology. Each family has the same structure, originated from the same ancestor, and has similar properties. Differently from the EC classification inside the family the substrate specificity and endo- or exo- mechanism can be different but the structure, catalytic amino acids and the catalytic mechanism is the same. Due to these, the determination of the structure of one member of

a family gives lot of information about all members of the investigated family (**Henrissat and Davies, 1997; Davies, 1998**).

Families can be further grouped into clans based on structure similarities. Each member of a clan has similar structure, and so the amino acids of the catalytic triad and catalytic mechanism are the same. Inside the clan, the substrate specificity can be different, but even enzymes belonging to different clans could have the same substrate specificity.

The available structures show that the mechanism of enzyme activity (endo/exo/disaccharide degradation) is not determined by the position of catalytic amino acids but by the structure of the substrate-binding region (**Fig. 2.1.3.1.1**). The catalytic amino acids are in a cleft in the case of endo acting enzymes, in a tunnel in the case of exo acting enzymes and in a pocket in the case of disaccharide degrading enzymes (**Aleshin et al. 1992; Spezio et al. 1993; Divne et al. 1994**).

The official name of the glycoside hydrolase contains the organism it is originated from, the most

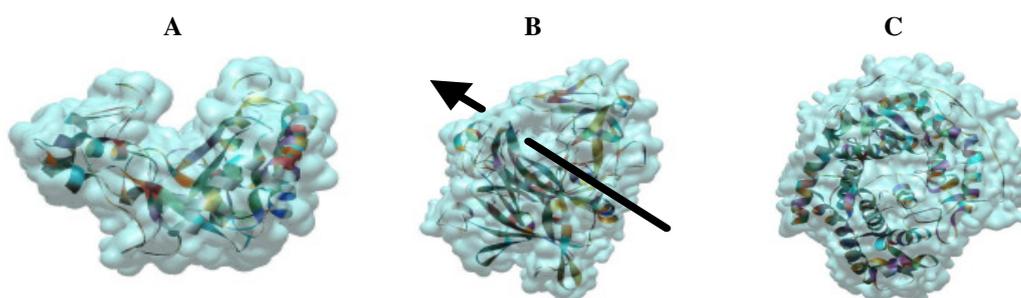


Fig. 2.1.3.1.1. Possible structure of the substrate-binding region. A: cleft ; B: tunnel ; C: pocket

preferred substrate, the number of the family it belongs and a letter showing the order of discovery. (PfXyn10A = *Pseudomonas fluorescens* firstly described family 10 xylanase) (**Henrissat et al. 1998**).

2.1.3.2. Carbohydrate binding modules (CBM)

Most of the non-catalytic modules present in glycoside hydrolase enzymes are carbohydrate binding modules, and very often cellulose binding modules. Almost all known water insoluble substrate degrading enzymes contain a substrate specific carbohydrate binding module which suggests that the carbohydrate binding modules are important in water insoluble substrate degradation (**Coutinho and Reilly, 1993; Blakk and Schrempf, 1995; Jespersen et al. 1991**).

The cellulose binding CBM are called cellulose binding modules (CBD). The xylane binding CBMs are called xylane binding modules (XBD) (**Black et al. 1995**). The starch binding CBMs are the starch binding modules (**Nunberg et al. 1984**).

In the carbohydrate-protein interactions the aromatic amino acids are extremely important. The ring of the tyrosine (Y), tryptophan (W) and phenylalanine (F) makes strong hydrophobic interaction (hydrophobic stacking) with the sugar ring, which is further enforced by H bonding of the sugar-OH

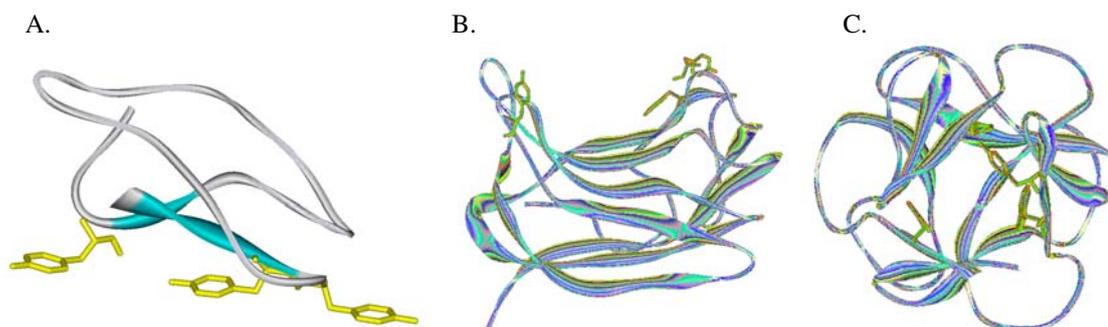


Fig. 2.1.3.2.1. Possible arrangement of substrate binding aromatic amino acids. A: along a plane: class A: CBM1 family (*Trichoderma reesei*, 1CBH.pdb); B: in a cleft: class B: CBM4 family (*Cellulomonas fimi*, 1ULO.pdb); C: in a packet: class C: CBM13 family (*Streptomyces olivaceoviridis*, 1XYF.pdb)

and amino acid-OH (Vyas, 1991).

Based on the structure, function and ligand binding ability the CBMs were categorized into three classes (Fig. 2.1.3.2.1.). CBMs belonging to class A are binding to the surface of water insoluble ligands, CBMs belonging to class B are binding to oligosaccharides longer than three sugar rings, and class C CBMs binds mono and disaccharides (Boraston et al. 1999).

2.1.4. Protein crystallization

During protein crystallization the aim is to produce such kind of supersaturated protein solution where the free energy loss happens by crystallization of the protein instead of its amorphous precipitation.

The perfectly ordered crystalline stage is the lowest free energy stage of the material, but it requires the formation of a higher energy stage intermedier during the nucleation. The amorphous form has a higher free energy stage than the crystalline stage but its formation does not require the higher energy stage intermedier. Because of this, the amorphous form is the preferred way of free energy decrease. Luckily, the conditions of oversaturation can be set in such a way that the nucleation can be started and crystals can form. Since the optimal conditions of nucleations are in a very narrow range, combination of several factors can only set it up. Due to this, very often the the crystallization solution contains at least 3 different components: the pH stabilizing **buffer**, the protein solubility determining **salt** and a **precipitator**. Different **additives** can further modify the effect of these components.

2.2 Materials and methods

Used strains:

E. coli BL21(DE3), BL21(DE3)*pLysS*, (*DE3*), *pLysS*, BMH 71-18 *mutS*, JM83, JM83(DE3), Tuner(DE3)TM, Oregami(DE3)TM, B834(DE3), JM101, TOP10

Used plasmids

pCRTMBlunt, pET16b, pET21a, pET21d, pGEX-4t-3

Used methods

Competent *E. coli* transformation, protein overexpression in *E. coli*; plasmid isolation from *E. coli*; PCR; site specific mutagenesis; periplasma isolation; His tagged protein isolation under native and denaturing conditions; Refolding of proteins expressed in inclusion bodies; Protein isolation by ion exchange chromatography and gelfiltration; protein crystallization; DNA and protein concentration determination; DNA separation in agarose gels; Protein electrophoresis SDS PAGE; DNA electrophoresis in agarose gels; DNA sequencing; Native PAGE; Quantitative avicell binding measurement; HPLC; ITC; CiDiSp

2.3. Results and discussion

Utilization of plant cell enzymes got very wide during the last two decades. It is used as much in food industry as for making more efficient washing powders or developing environment friendly methods on different industrial fields and in the bio-ethanol production. The understanding of the enzyme action is preliminary requirement to fulfill the increasing demands of generating enzymes that are more special either by making the enzymes more active, more specific, or more tolerant to the environment.

Plant cell wall enzymes very often contain substrate-binding modules that are important for the proper enzyme activity against insoluble ligands. The understanding of the mechanism of ligand binding could help to produce very effective enzymes for bio-fuel production.

Aromatic amino acids on the surface of the substrate binding modules are the keys in the ligand binding, where strong hydrophobic stacking reactions stabilize the position of the ligand in the proteins' substrate binding cleft.

Aromatic amino acids involved in the insoluble cellulose binding were identified by site-specific mutagenesis in the case of CBM10. The five conserved aromatic amino acids of the CBM10 family

were replaced with alanine respectively. While mutations W7A, W22A, W24A and Y8A abolished, the Y2A mutation did not influence the cellulose binding ability of the catalytic module fused to family 10 CBD. Although no significant difference was seen in the CiDi (Circular dichroism) spectra of the wild type and

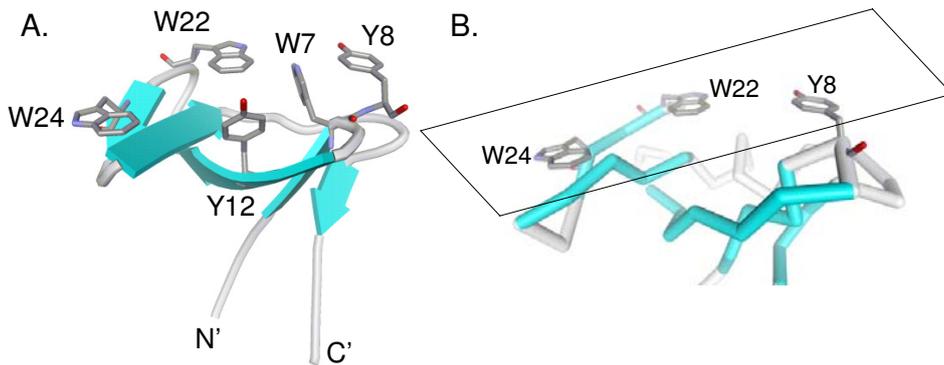


Fig. 2.3.1. Structure of the CBM10 module determined by NMR. A: position of the conserved amino acids B: Trp22, Trp24 and Tyr8 aminoacids arranged along the plane

mutant proteins, the standalone W7A CBM could not be expressed in *E. coli* which suggested that the W7 is required for maintaining the correct 3D structure of the module. NBS (N-bromo-succinamide) was used to determine the number of surface tryptophans potentially participating in the substrate binding. Under denaturing conditions three, under native conditions, only two tryptophans reacted with NBS of the wild type CBM10 module. In the case of W22A and W24A mutants under native conditions, only one tryptophan showed reaction with NBS. Based on these results out of the five conserved aromatic amino acids of the CBM10 family only Y8, W22 and W24 participates in the substrate binding since the replacement of Y12 with alanine did not influence the substrate binding and the W7 was shown not be on the surface of the CBM10 module. These results were confirmed by the NMR structure of the wild type CBM10. The rings of the aromatic amino acids of W22, W24 and Y8 are arranged along a plane (**Fig. 2.3.1.**) enabling the binding to the similarly arranged glucose rings of cellulose.

The other way to identify amino acids participating in ligand binding is to determine the 3D structure of the binding module. Unfortunately, the preliminary requirement of structure determination is the calculation of initial phases. This is mostly done by collecting diffractive data with the native crystal and crystals containing heavy metal ions. Luckily, the initial diffractive phases can be determined without heavy metal ions. It could be done either by replacing the sulphurs with selenium and the diffraction data are collected at three different wavelengths

(sometimes the larger selenium could inhibit crystallization of the protein under the same conditions). The other possibility is to crystallize the target protein when it is non-flexibly attached to a protein with known structure. Using the known structure the structure of the non-flexible fusion protein can be determined using molecular replacement. This method eliminates the need of heavy metal ions or selenomethionine. Molecular replacement can be used even when that part of the fusion protein, which is used for initial phase calculations, has high amino acid similarity to known structures.

The PsXyn10F catalytic module shows 60 % amino acid identity with the catalytic domain of PsXyn10A with known 3D structure. The

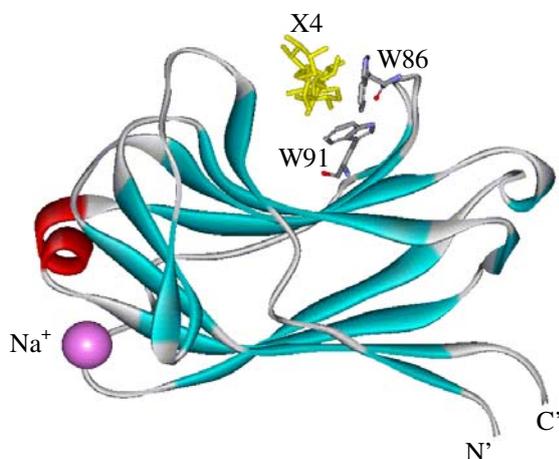


Fig. 2.3.2. Structure of the CBM15 module in the presence of 15 mM xylopentose. In the binding cleft the xylopentose (X4) and the binding aromatic amino acids (W86 and W91) can be seen

CBM15 xylan binding module is linked to the PsXyn10F catalytic domain without obvious linker region, which could mean that the substrate binding cleft of the CBM15 module is the continuation of the substrate binding cleft of the catalytic domain. If it is true than the two modules (catalytic and xylane binding) must have a rigid connection and the structure of the xylan binding module could be determined using molecular replacement starting from the catalytic module. To test this, the catalytic module was expressed together with the xylan binding module. The purified protein was crystallized and molecular replacement was used to determine first the structure of the catalytic module. Unfortunately using the structure information of the catalytic domain only 30 % of the CBM15 structure could be determined, since the position of the xylan binding module was not well defined in the crystal. This showed that the connection between the catalytic module and the xylan binding module is not rigid. Luckily, the determined structure of the CBM15 module was enough to determine the structure of the standalone crystallized CBM15 module in the presence of xylopentose (**Fig. 2.3.2**). The obtained structure was the first experimental prove that the xylan has 180° turning structure. Using the CBM15 structure, the aromatic amino acids important in xylan binding were determined, and we were able to explain why this module is able to bind substituted xylans.

During the above mentioned work, the structure of the PsXyn10F catalytic domain was determined

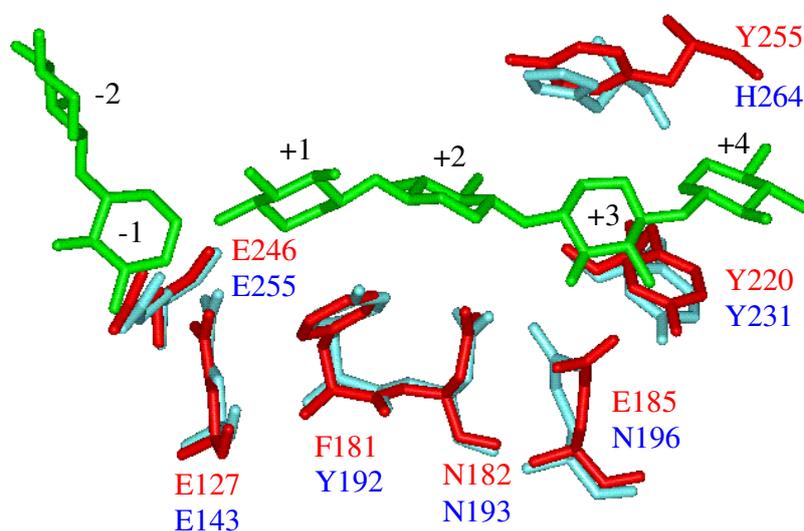


Fig. 2.3.3. Catalytic and substrate binding amino acids of the XynA and XynF catalytic domains

and compared with the PsXyn10A catalytic domain which structure was already determined in the presence of different substrates. After aligning the structures, the amino acids important in substrate binding and hydrolysis were identified (**Fig. 2.3.3**). The only significant difference was that the Glu43 amino acid of the XylA protein was replaced with the small glycine in XylF. Earlier works showed that when the XylA Glu43 was replaced with alanine, the

enzyme activity of the mutated XylA was significantly reduced against PNPC (p-Nytrufenil-β-D-cellobiose). Based on this XylF must have low PNPC activity compared to XylA. The comparison of the PNPC activity of the XylA and XylF proved the assumption.

Sometimes the protein under investigation does not show any homology to characterized structures. Under these conditions the easiest way of understanding how the protein fulfills its function is to determine its structure in the presence of its substrate.

The structure of the multiple soluble-substrate binding CBM29-2 was determined in the presence of mannohexose and cellobiose (**Fig. 2.3.4**). Based on the binding between the glucose and mannose units and substrate binding amino acids, the structure of the theoretical most proper substrate was forecasted. The structure of the theoretical best substrate was similar to the ITC experiments determined substrate which was the konjack glucomannan. Among the tested homopolymers, mannopentose showed the weakest binding affinity, and cellopentose, cellobiose and HEC showed higher and higher binding affinity depending on the number of the sugar units.

The 37% homologous CBM29-1 module shows much lower affinity to the tested substrates than the CBM-29-2. Using molecular modeling it was shown that the ligand binding cleft contains several huge amino acids which could be the reason of the lower substrate affinity of the CBM-29-1.

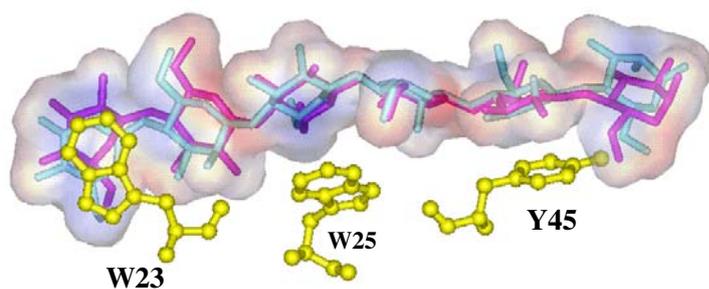


Fig. 2.3.4. The position of the celohexose and mannohexose chains compared to the position of the substrate binding aromatic amino acids of the CBM-29-2 module Purple: celohexose; Blue: mannohexose, Yellow: substrate binding aromatic amino acids

The limiting step in structure determination is the protein crystallization. Very often long and painful way leads to the identification of the optimal conditions of protein expression, purification and crystal formation to get the protein crystal suitable for structure determination.

The X4 module of PfMan10 enzyme was tried to be crystallized. The X4 module was expressed with N' or C' terminal His tag and without His tag. It was isolated from the cytoplasm or from the periplasm. It was crystallized in the presence and absence of several different substrates. After a long hassle, 1 M CaCl₂ and 10 % DMF resulted diffracting crystals.

Summary of the new scientific results

- an essential enzyme coding ORF was isolated from the bacitracin producer *Bacillus licheniformis* strain which was named *btsT* due to its homology to thioesterase II proteins;
 - it was proved that the inactivation of the *btsT* ORF severely reduced the bacitracin production;
 - it was proved that the copy number increase of the *btsT* gene reduces mostly the production of the bacitracin B₁ and bacitracin B₂ components without significantly altering the bacitracin A component production. Based on this, thioesterase genes could be important tools of impurity decrease in industrial applications;
 - based on other homologue proteins, the fold of the BtsT protein was determined. It belongs to the α - β folded protein group and haloperoxidase subgroup;
 - using the haloperoxidase structure the members of the catalytic triad were identified
 - a β -galactosidase based reporter gene system was created to determine the time course of the expression of the *btsT* gene and the bacitracin synthetase first subunit coding gene;
 - based on the measurement it was concluded that the transcription from the bacitracin synthetase first subunit coding gene is much higher than that of the thioesterase gene. The bacitracin production starts when both genes reach maximal transcription.
-
- Substrate binding aromatic amino acids of the CBM10 cellulose binding module originated from *Pseudomonas fluorescens* Xyn10A were determined by site specific mutagenesis and NBS titration of the surface tryptophans.
 - The structure of the xylan binding CBM15 module originated from *Pseudomonas fluorescens* XylF was determined in the presence of xylohexose.
 - The structure of water soluble substrate binding CBM29-2 module originated from *Piromyces equi* CelC was determined in the presence of xylohexose and mannohexose. Based on the structure information the composition of the ideal substrate was described.

Used literature

1. Aharonowitz, Y., Bergmeyer, J., Cantoral, J. M., Cohen, G., Demain, A. L., Fink, U., Kinghorn, J., Kleinkauf, H., Mac-Cabe, A., Palissa, H., et al. (1993) *BioTechnology* 11, 807-810.
2. Albersheim, P., An, J., Freshour, G., Fuller, M.S., Guillen, R., Ham, K.S., Hahn, M.G., Huang, J., O'Neill, M., Whitcombe, A., Williams, M.V., York, W.S. and Darvill, A. (1994). Structure and function studies of plant cell wall polysaccharides. *Biochem. Soc. Trans.* 22: 374-378
3. Aleshin, A., Golubev, A., Firsov, L. M. and B., H. R. (1992). Crystal structure of glucoamylase from *Aspergillus awamori* var. X100 to 2.2 Å resolution. *J Biol Chem* 267: 19291-19298.
4. Biely P, Mislovicova D, Toman R. (1985) Soluble chromogenic substrates for the assay of endo-1,4-beta-xylanases and endo-1,4-beta-glucanases. *Anal Biochem.* 144(1):142-6.
5. Black, G.W., Hazlewood, G.P., Millward-Sadler, S.J., Laurie, J.I. and Gilbert, H.J. (1995). A modular xylanase containing a novel non-catalytic xylan-specific binding domain. *Biochem. J.* 307: 191-195
6. Boraston, A.B., Mclean, B.W., Kormos, J.M., Alam, M., Gilkes, N.R., Haynes, C.A., Tomme, P., Kilburn, D.G. and Warren, R.A.J. (1999). Carbohydrate binding modules: diversity of structure and function. 202-211. in Gilbert, H.J., Davies, G.J., Henrissat, B. and Svensson, B. (eds.), *Recent Advances in Carbohydrate Bioengineering*. The Royal Society of Chemistry, Cambridge.
7. Brett, C. and Waldren, K. (1996). *Physiology and Biochemistry of Plant cell walls*. Topics in plant functional biology: 1. M. Black and B. Charlewood (eds). Chapman and Hall.
8. Brett, C. T. and Waldren, K. (1990). *Physiology and biochemistry of plant cell walls*. Topics in plant physiology: 2. M. Black and J. Chapman (eds.). London, Unwin Hyman.
9. Cane, D.E and Walsh, C.T. (1999). The parallel and convergent universes of polyketide synthases and nonribosomal peptide synthetases. *Chem. Biol.* 6: R319-R325
10. Challis G.L., Ravel J. and Townsend C.A. (2000) Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chem. Biol.* 7 211- 224
11. Coughlan, M.P. (1985). Enzymic hydrolysis of cellulose: an overview. *Biores. Technol.* 39:107- 3: 853-859.
12. Davies, G.J. (1998). Structural studies on cellulases. *Biochem. Soc. Trans.* 26: 167-172.
13. Dekker R.F.H. and Richards, D.N. (1976). Hemicellulases: their occurrence, purification, properties and mode of action. *Adv. Carb. Chem. Biochem.* 32: 261-264
14. Dey, P.M. and Brinson, K. (1984). Plant cell walls. *Adv. Carb. Chem. Biochem.* 42: 265-382.
15. Divne, C., Stahlberg, J., Reinikainen, T., Ruohonen, L., Petterson, G., Knowles, J. K. C., Teeri, T. T. and Jones, T. A. (1994). The three dimensional crystal structure of the catalytic core of cellobiohydrolase I from *Trichoderma reesei*. *Science* 265: 524-528.
16. reesei. *Science* 265: 524-528.
17. Doekel, S. and Marahiel, M.A. (2000). Dipeptide formation on engineered hybrid peptide synthetases. *Chemistry & Biology*, 7:373-384
18. Epperson J.D. and Ming L.J.(2000) Proton NMR Studies of Co(II) Complexes of the Peptide Antibiotic Bacitracin and Analogues: Insight into Structure-Activity Relationship *Biochemistry*, 39, 4037-4045
19. Gilkes, N. R., Kilburn, D. G., Miller, R. C. and Warren, R. A. J. (1988). Cellulases of *Cellulomonas fimi* - the enzymes and their interaction with substrate. *Abst Papers Am Chem Soc* 195: 123.
20. Hazlewood GP and Gilbert HJ. (1998b) *Pseudomonas* cellulose-binding domains mediate their effects by increasing enzyme substrate proximity. *Biochem J:* 331 (Pt 3):775-81.
21. Hazlewood, G. P. and Gilbert, H. J. (1998a). Structure and function analysis of *Pseudomonas* plant cell wall hydrolases. *Biochem. Soc. Trans.* 26, 185-190.
22. Heikinhemino, P., Goldman, A., Jeffris, C. and Ollis, D.L. (1999). Of barn owls and bankers: a lush variety of α/β hydrolases. *Structure* 7: R141-R146
23. Henrissat, B. and Coutinho, P. (2000). CAZy website: <http://afmb.cnrs-mrs.fr/~pedro/CAZY/db.html>.
24. Henrissat, B. and Davies, G. (1997). Structural and sequence based classification of glycoside hydrolases. *Curr. Op. Struct. Biol.* 7: 637-644.
25. Henrissat, B., Teeri, T. T. and Warren, R. A. J. (1998). A scheme for designating enzymes that hydrolyse the polysaccharides in the cell walls of plants. *FEBS Lett.* 425 352-354.
26. Hughey, R., A. Krogh, (1996) Hidden Markov models for sequence analysis: Extension and analysis of the basic method, *CABIOS* 12:95-107.
27. Ikai, Y., Oka, H., Hayakawa, J., Matsumoto, M., Saito, M., Harada, K., Mayumi, T. and Suzuki, M. (1994). Total structures and antimicrobial activity of bacitracin minor components. *J. Antibiot.* 48 (3) 233-242.
28. Joseleau, J.P., Comtat, J. and Ruel, K. (1992). Chemical structure of xylans and their interactions in the plant cell walls. In: *progress in Biotechnology, Vol 7: Xylans and Xylanases*, 1-15, J. Visser, G. Beldman, M.A.K.-v. Someren and A.G.J. Voragen (eds.). Elsevier, Amsterdam.
29. Karplus, K., C. Barrett, R. Hughey (1999) Hidden Markov models for detecting remote protein homologies, *Bioinformatics* 14(10):846-856
30. Knox, J.P., Linstead, P.J., King, J., Cooper, C. and Roberts, K. (1990). Pectin esterification is spatially regulated both within cell walls and between developing tissues of root apices. *Planta* 181: 512-521

31. Konz D, Klens A., Schorgendorfer K. and Marahiel M.A. (1997) The bacitracin biosynthesis operon of *Bacillus licheniformis* ATCC 10716: molecular characterization of three multi-modular peptide synthetases *Chem. Biol.* 4 (12), 927-937
32. Kormelink, F. J. M., Gruppen, H. and Voragen, A. G. J. (1993). Mode of action of (1,4)- β -D-arabinoxylan arabinofuranohydrolase (AXH) and α -L-arabinofuranosidases on alkali-extractable Wheat-flour arabinoxylan. *Carbohydr. Res.* 249: 345-353.
33. Krogh A, Brown M, Mian IS, Sjolander K, Haussler D. (1994) Hidden Markov models in computational biology: Applications to protein modeling, *JMB* 235: 1501-1531.
34. Laland, S.G. and Zimmer, T. (1973). The protein template mechanism of synthesis for the peptide antibiotics produced by *Bacillus licheniformis* *Essays Biochem.* 9. 31-35
35. Lambalot RH, Gehring AM, Flugel RS, Zuber P, LaCelle M, Marahiel MA, Reid R, Khosla C, Walsh CT. (1996). A new enzyme superfamily — the phosphopantetheinyl transferases. *Chem. Biol.* 3, 923–936.
36. Libertini, L. J. and Smith, S. (1978) Purification and properties of a thioesterase from lactating rat mammary gland which modifies the product specificity of fatty acid synthetase. *J. Biol. Chem.* 253: 1393-1401
37. Linne, U., and Marahiel, M.A. (2000). Control of directionality in nonribosomal peptide synthesis: role of the condensation domain in preventing misinitiation and timing of epimerization. *Biochemistry* 39, 10439–10447.
38. Marahiel, M.A., Stachelhaus, T. and Mootz, H.D. (1997). Modular peptide synthetases involved in non-ribosomal peptide synthesis. *Chem. Rev.* 97, 2651-2673.
39. Morris, M. (1994) Primary structural confirmation of components of the bacitracin complex *Biol. Mass Spectrosc.* 23, 61-70.
40. Neumuller AM, Konz D, Marahiel MA. (2001) The two-component regulatory system BacRS is associated with bacitracin 'self-resistance' of *Bacillus licheniformis* ATCC 10716. *Eur J Biochem.* 268 (11): 3180-9
41. Nicholson WL, Setlow P. (1990) Dramatic increase in negative superhelicity of plasmid DNA in the forespore compartment of sporulating cells of *Bacillus subtilis*. *J. Bacteriol.* 172 (1): 7-14.
42. Nunberg J.H., Meade J.H., Cole G., Lawyer F.C., McCabe P., Schweickart V., Tal R., Wittman V.P., Flatgaard J.E., Innis M.A. (1984) Molecular cloning and characterization of the glucoamylase gene of *Aspergillus awamori*. *Mol. Cell. Biol.* 4: 2306-2315.
43. Pazirandeh, M., Chirala, S.S. and Wakil, S.J. (1991) Site-directed mutagenesis studies on the recombinant thioesterase domain of chicken fatty acid synthase expressed in *Escherichia coli* *J. Biol. Chem.* 266: 20946-20952.
44. Pazirandeh, M., Chirala, S.S., Huang, W.Y. and Wakil, S.J. (1989) Characterization of recombinant thioesterase and acyl carrier protein domains of chicken fatty acid synthase expressed in *Escherichia coli*. *J. Biol. Chem.* 264: 18195-18201.
45. Podlasek Z, Comino,A., Herzog-Velikonja,B., Zgur-Bertok,D., Komel,R. and Grabnar,M. (1995) *Bacillus licheniformis* bacitracin-resistance ABC transporter: relationship to mammalian multidrug resistance *Mol. Microbiol.* 16 (5), 969-976
46. Poutanen, K. (1988). An α -L-arabinofuranosidase of *T. reesei*. *J. Biotechnol.* 7: 271-282
47. Prágai Z., Holczinger, A. and Sík, T. (1994/a) Transformation of *Bacillus licheniformis* protoplasts by plasmid DNA *Microbiology*, 140: 305-310
48. Prágai, Z., Tran, S.L.P., Nagy, T. Fülöp, L., Holczinger, A. and Sík, T. (1994/b). Transposon Tn917PF1 mutagenesis in *Bacillus licheniformis* *Microbiology* 140, 3091-3097
49. Schneider, A. and Marahiel, M.A. (1998). Genetic evidence for a role of thioesterase domains, integrated in or associated with peptide synthetases, in non-ribosomal peptide biosynthesis in *Bacillus subtilis*. *Arch. Microbiol.* 169, 404-410.
50. Shaw-Reid, C. A., Kelleher, N. L., Losey, H. C., Gehring, A M., Berg, C., and Walsh, C. T. (1999) *Chem. Biol.* 6, 385-400.
51. Siegel, M. M., Huang, J., Lin, B., and Tsao, R. (1994) Structures of bacitracin A and isolated congeners: sequencing of cyclic peptides with blocked linear side chains by electrospray ionization mass spectrometry. *Biol. Mass Spectrosc.* 23, 196-204.
52. Spezio, M., Wilson, D. B. and Karplus, P. A. (1993). Crystal structure of the catalytic domain of a thermophilic endocellulase. *Biochem.* 32: 9906-9916.
53. Stachelhaus, T. and Marahiel, M. A. (1995/a). Modular structure of genes encoding multifunctional peptide synthetases required for non-ribosomal peptide synthesis *FEMS Microbiol. Lett.* 125, 3-14
54. Stachelhaus, T. and Marahiel, M. A. (1995/b). Modular structure of peptide synthetase revealed by dissection of the multifunctional enzyme GrsA. *Journal of Biological Chemistry* 270: 6163-6169
55. Stachelhaus, T., Mootz, H.D., Bergendahl, V. and Marahiel, M.A. (1998). Peptide bond formation in nonribosomal peptide biosynthesis. Catalytic role of the condensation domain. *J. Biol. Chem.* 273, 22773-22781.
56. Stachelhaus, T., Mootz, H.D., Marahiel, M.A. (1999) The specificity-conferring code of adenylation domains in non-ribosomal peptide synthetases. *Chem. Biol.* 6(8) 493- 505
57. Stein, T., Vater, J., Kruft, V., Otto, A., Wittmann-Liebold, B., Franke, P., Panico, M., McDowell, R., and Morris, H. R. (1996). The multiple carrier model of nonribosomal peptide biosynthesis at modular multienzymatic templates. *J. Biol. Chem.* 271, 15428-15435.

58. Tai, M.H., Subrahmanyam, S.C., and Wakil, S.J. (1993). Roles of Ser101, Asp236 and His237 in catalysis of thioesterase II and the C terminal reegion of the enzyme in its inetraction with fatty acid synthase. *Proc. Natl. Acad. Sci. USA* 90: 1852-1856.
 59. Tomme, P. Van Tilbeurgh, H., Petterson, G., Van Damme, J., Vandekerckhove, J., Knowles, J., Teeri, T. and Claeysens, M. (1988). Studies of the cellulolytic system of *Trichoderma reesei* QM9414. Analysis of domain finction in two cellobiohydrolases by limited proteolysis. *Eur. J. Biochem.* 170: 575-581
 60. Tomme, P., Warren, R. A. J. and Gilkes, N. R. (1995). Cellulose hydrolysis by bacteria and fungi. *Adv. Micro. Physiol.* 37: 1-81.
 61. van Wageningen A.M., Kirkpatrick, P., Williams, D., Harris, B., Kershaw, J., Lennard, N., Jones, M., Jones, S., and Solenberg, P. (1998) Sequencing and analysis of genes involved in the biosynthesis of a vancomycin group antibiotic. *Chem. Biol.* 5, 155-162.
 62. Varner, J. E. and Lin, L. S. (1989). Plant cell wall architecture. *Cell* 56: 231-239.
 63. Vyas, N.K. (1991). Atomic features of protein carbohidrate interactions. *Curr. Opin. Struct. Biol.* 1: 732-740.
 64. Wakil, S.J. (1989). Fatty acid synthase, a proficient multifunctional enzyme. *Biochemistry* 28, 4523-4530.
 65. Warren, R.A.J. (1996). Microbial hydrolysis of polysaccharides. *Ann. Rev. Microbiol.* 50: 183-212.
 66. Yui, T., Miyawaki, K., Yada, M. and Ogawa, K. (1997). An evaluation of crystal structure of mannan I by X-ray powder diffraction and molecular mechanics studies. *Internat. J. Biol. Macromol.* 21, 243-250.
 67. Zocher, R., Nihira, T., Paul, E., Madry, N., Peeters, H., Kleinkauf, H., and Keller, U. (1986) Biosynthesis of cyclosporin A: partial purification and properties of a multifunctional enzyme from *Tolypocladium inflatum*. *Biochemistry* 25, 550-553.
-

Publications directly related to the thesis

Szabo L, Jamal S, Xie H, Charnock SJ, Bolam DN, Gilbert HJ, Davies GJ. (2001) Structure of a family 15 carbohydrate-binding module in complex with xylopentaose. Evidence that xylan binds in an approximate 3-fold helical conformation. *J Biol Chem.* 276(52):49061-5. (IF: 7,258)

Ponyi T, **Szabo L**, Nagy T, Orosz L, Simpson PJ, Williamson MP, Gilbert HJ. (2000) Trp22, Trp24, and Tyr8 play a pivotal role in the binding of the family 10 cellulose-binding module from *Pseudomonas xylanase A* to insoluble ligands. *Biochemistry.* 39(5):985-91. (IF: 4,221)

Raghothama S, Simpson PJ, **Szabo L**, Nagy T, Gilbert HJ, Williamson MP. (2000) Solution structure of the CBM10 cellulose binding module from *Pseudomonas xylanase A*. *Biochemistry.* 39 (5): 978-84. (IF: 4,221)

Charnock SJ, Bolam DN, Nurizzo D, **Szabo L**, McKie VA, Gilbert HJ, Davies GJ. (2002) Promiscuity in ligand-binding: The three-dimensional structure of a *Piromyces* carbohydrate-binding module, CBM29-2, in complex with cello- and mannohexaose. *Proc Natl Acad Sci USA.* 99(22):14077-82. (IF: 10,896)

Pell G, **Szabo L**, Charnock SJ, Xie H, Gloster TM, Davies GJ, Gilbert HJ. (2004) Structural and biochemical analysis of *Cellvibrio japonicus* xylanase 10C: how variation in substrate-binding cleft influences the catalytic profile of family GH-10 xylanases. *J Biol Chem.* 279 (12): 11777-88. (IF: 6,335)

Other publications

Jahn M, Stoll D, Warren RA, **Szabo L**, Singh P, Gilbert HJ, Ducros VM, Davies GJ, Withers SG. (2003) Expansion of the glycosynthase repertoire to produce defined manno-oligosaccharides. *Chem Commun (Camb).* (12): 1327-9. (IF: 4,031)

Ducros VM, Zechel DL, Murshudov GN, Gilbert HJ, **Szabo L**, Stoll D, Withers SG, Davies GJ. (2002) Substrate distortion by a beta-mannanase: snapshots of the Michaelis and covalent-intermediate complexes suggest a B(2,5) conformation for the transition state. *Angew Chem Int Ed Engl.* 41 (15): 2824-7. (IF: 8,255)

Xie H, Bolam DN, Nagy T, **Szabo L**, Cooper A, Simpson PJ, Lakey JH, Williamson MP, Gilbert HJ. (2001) Role of hydrogen bonding in the interaction between a xylan binding module and xylan. *Biochemistry.* 40 (19): 5700-7. (IF: 4,114)

Tran LS, **Szabo L**, Ponyi T, Orosz L, Sik T, Holczinger A. (1999) Phage abortive infection of *Bacillus licheniformis* ATCC 9800; identification of the abiBL11 gene and localisation and sequencing of its promoter region. *Appl Microbiol Biotechnol.* 52 (6): 845-52. (IF: 1,641)

Tran LS, **Szabo L**, Orosz L, Sik T, Holczinger A. (1998) Construction of a single-copy integration vector and its use to study gene expression in *Bacillus licheniformis*. *Microbiology.* 144 (Pt 9): 2573-8. (IF: 2,564)

Tran LS, **Szabo L**, Fulop L, Orosz L, Sik T, Holczinger A. (1998) Isolation of a beta-galactosidase-encoding gene from *Bacillus licheniformis*: purification and characterization of the recombinant enzyme expressed in *Escherichia coli*. *Curr Microbiol.* 37 (1): 39-43. (IF: 1,094)

Szabo L and Tchelet R. (2006) Process for constructing strain having compactin hydroxylation ability United States Patent Application 20060172383 (Filed: December 5, 2005)