PhD thesis

Regulation of nopaline uptake in *Agrobacterium tumefaciens*

Dr Ferenc Marincs

Agricultural Biotechnology Center
Gödöllő

2007

Contents

Foreword

Agrobacterium is mainly known amongst scientists because certain virulent strains of this bacterium, which harbour large plasmids, are able to transfer particular plasmid genes into the genome of a wide variety of plants. This characteristic is very important in plant molecular biology, where it is widely used to produce transgenic plants for basic research and agriculture. Therefore, the main focus of *Agrobacterium* research was and still is this gene transfer or transformation as it is usually called. There are a large number of publications which deal with issues regarding *Agrobacterium*-mediated transformation of plants, for example which proteins and DNA structures are involved in the transformation process; how this process is regulated; how the transferred genes are expressed once integrated into the plant genome and how transgenic plants are produced using the transformation method.

In nature, the transformed plants cells provide an environment where the bacteria can survive on unique substances, collectively called opines, which are produced by the transformed plant cells and can be utilised as a sole carbon and nitrogen source by the colonising bacteria. Genes encoding the catabolic proteins for these opines are encoded on the same large plasmids, a particular portion of which is transferred into the plant genome causing tumour formation and opine synthesis.

I applied for and received a research position in New Zealand in 1989, to study the molecular biology of opine catabolism. Although the “opine concept” describing the role of opines in the tumour-*Agrobacterium* interaction, was described a long time ago, not much was known about opine catabolism at the molecular level at that time. In the laboratory, which I joined in 1989, Dr Derek White had previously mapped the nopaline catabolism genes of the virulent *Agrobacterium tumefaciens* Ti plasmid pTiT37 by transposon mutagenesis, and I continued his work to study gene regulation in the nopaline catabolism operon. The goals of my work were to find and characterise those elements which are involved in the regulation of nopaline catabolism of the Ti plasmid pTiT37 and to learn the way in which these elements regulate gene expression. This thesis describes my work carried out between 1989 and 1995 and based on four first-authorship papers published between 1993 and 1996.

Where appropriate, I have used recent publications to update information about the topic.

Abbreviations throughout the text are explained at the first occurrence.

Acknowledgement

First, I would like to thank Dr Derek White who trusted and hired me to do the work described in this thesis in his group at AgResearch (formerly DSIR) in Palmerston North, New Zealand, where I spent eleven unforgettable years. In particular I thank him for giving me a free hand to design and conduct all experimental work, draw the conclusions and write publications about the topic.

I could not perform such a work without a solid background in genetics and molecular biology, which I gained at the Department of Genetics, Szeged University in the era of Professor László Orosz, where I had the opportunity to do research and teaching for ten years, and also to receive my university doctorate degree in 1986. I am very grateful to Professor Orosz for this opportunity and I am very glad that I was part of that fantastic community.

I learned the basic techniques of molecular biology at the Boyce Thompson Institute, Cornell University, Ithaca, USA, in the laboratory of Professor Aladár Szalay, and I am thankful for this.

I also thank those people in all three labs, who helped and supported me, or simply became good friends and colleagues. All of them have some, even tiny, professional or human contribution to this thesis, but there are so many of them that I have to apologise not to list everyone's name.

I thank my late parents, who supported me through my studies and private life and without their loving care and help I would not be who I am now; and my wife who always stood on my side even in the most difficult times.

Finally I thank Professor Simon Baumberg, Leeds University for his critical reading of the manuscript.

1. Introduction

1.1. Regulation of gene expression in bacteria

1.1.1. General issues

In every organism, including bacteria, transcription is the first step of processes that provide gene products for cellular biochemistry. In order to ensure coordinated spatial and temporal presence of these products, transcription must be carefully regulated. Fidelity of control is achieved by employing regulatory proteins, which interact with specific binding sites called operator sequences (or in short: operators) in or around the promoter region of the regulated genes, and either prevent or promote the RNA polymerase-mediated transcription of the genes. Regulatory proteins can be categorised according to their regulatory function. They can be either repressors, which inhibit, or activators, which facilitate transcription by binding to their operator sites. There are also regulatory proteins with dual nature whose repressor or activator function depends on the conditions. In *Escherichia coli*, amongst 314 characterised DNA-binding regulatory proteins, 43% were repressors, 35% activators and 22% dual regulators (Pérez-Rueda and Collado-Vides 2000).

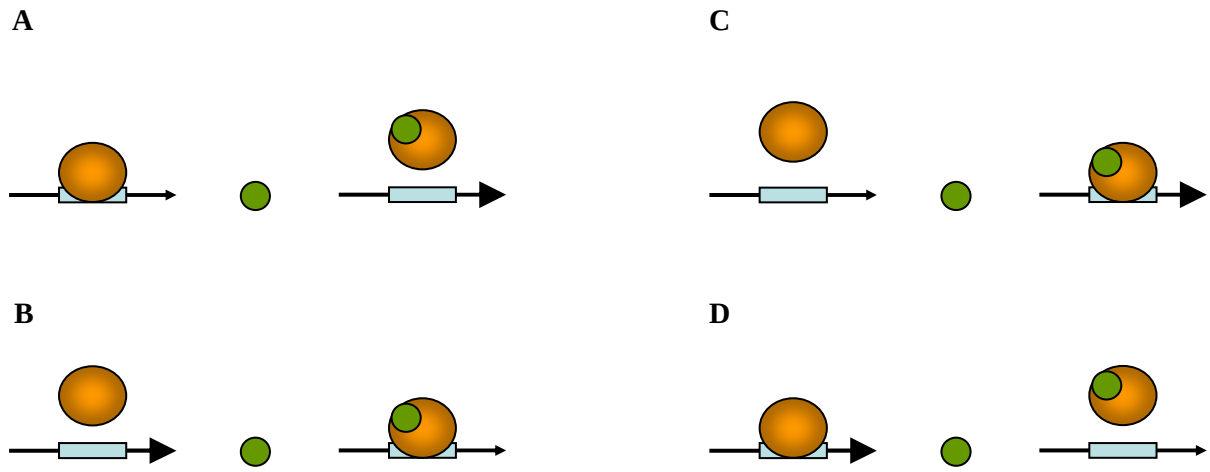

Figure 1: Repression (A and B) and activation (C and D) by regulatory proteins and small mediator molecules (Lewin 1990). Blue boxes, yellow and green circles represent operators, regulatory proteins and small mediator molecules, respectively. The size of the arrowheads indicates the level of gene expression.

Both repression and activation can be mediated by small molecules (inducers, co-repressors or co-activators). Repressors and activators are released from or attached to their operators in the

presence or absence of these small mediator molecules in order to accomplish their function (Figure 1).

Small effector molecules have their effect on regulatory proteins by influencing either the protein's affinity for the operator or its conformation or both (Weickert and Adhya 1992), and consequently, the function of the regulatory protein is affected. In dual regulators, binding of the effector molecule causes a conformational change in the protein, but usually does not influence its DNA-binding affinity. That conformational change acts as a switch between the protein's repressing and activating functions (Wang et al. 1992).

1.1.2. Regulatory proteins

In general, bacterial regulatory proteins contain an even number of the same subunits. Two or four subunits are the most common (Weickert and Adhya 1992), but there are examples for regulatory proteins containing six subunits (Holtham et al. 1999). Subunits of regulatory proteins have different functional domains; about three-quarters of them have been identified as two-domain proteins (Babu and Teichmann 2003). The DNA-binding domain, as indicated by its name, is responsible for the interaction between the regulatory protein and the operator, while other domains contain sequences for subunit interaction and binding effector molecules. These functional domains can be positioned anywhere along the subunit, although DNA-binding domains are usually located in the N-terminal while other domains are found in the C-terminal of regulatory proteins (Pérez-Rueda and Collado-Vides 2000). A domain either displays an independent function or contributes to the function of a multidomain protein in interaction with other domains (Vogel et al. 2004). Chimeric proteins constructed from domains of related proteins often demonstrate domain independency (Ladant and Karimova 2000).

The DNA-binding domain contains a sequence motif, which forms a special tertiary structure that binds to the operator sequence. The most common is the helix-turn-helix (HTH) motif (Harrison 1991; Pérez-Rueda and Collado-Vides 2000, 2001), which is approximately 20 residues long and characterised by two alpha-helices connected by a short turn (Brennan and Matthews 1989). While the first helix stabilises the structure, the second helix of the HTH motif binds to DNA (Figure 2) via hydrogen-bonds and hydrophobic interactions which occur between side chains of the helix and groups in the major groove of the DNA.

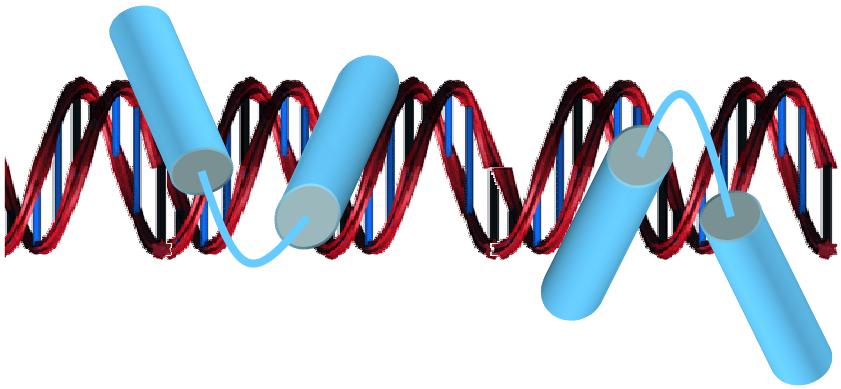

Figure 2: Structure of the HTH-DNA complex. One helix (blue cylinder) of each motif binds in the major groove of the DNA.

The position of the motif is characteristic for different type of regulatory proteins. Repressor proteins tend to have the HTH motif in their N-terminus, while activators are the opposite: the majority of them have the HTH motif in the C-terminus. Dual function regulatory proteins predominantly have the HTH motif in their N-terminus (Pérez-Rueda and Collado-Vides 2000).

In a minor proportion of bacterial regulatory proteins, other DNA-binding motifs, such as zinc-finger, helix-loop-helix (HLH) and ribbon-helix-helix (RHH) domains have been identified (Pérez-Rueda and Collado-Vides 2000).

1.1.3. Binding sites of regulatory proteins

In the sequences to which regulatory proteins bind, there are short (usually less than 20 bp), highly symmetrical DNA sequences, called operators, to which regulatory proteins bind in order to mediate their regulatory functions. The symmetrical part of the operator sequence can be a palindrome or an inverted or tandem repeat or a combination of these (Figure 3).

The position of an operator relative to the regulated promoter is important in terms of the mechanism of the regulation. In most promoters the operator site is located between the -60 and +20 positions, i.e. their position is termed proximal. This type of operators is quite common and can be found in a large number of promoters. In almost half of the repressible promoters, the operator is positioned around the +1 position, while in 60 % of the activable promoters the operator overlaps the -40 position (Collado-Vides et al. 1991). The number of proximal operators can vary between one and five in the regulated promoters.

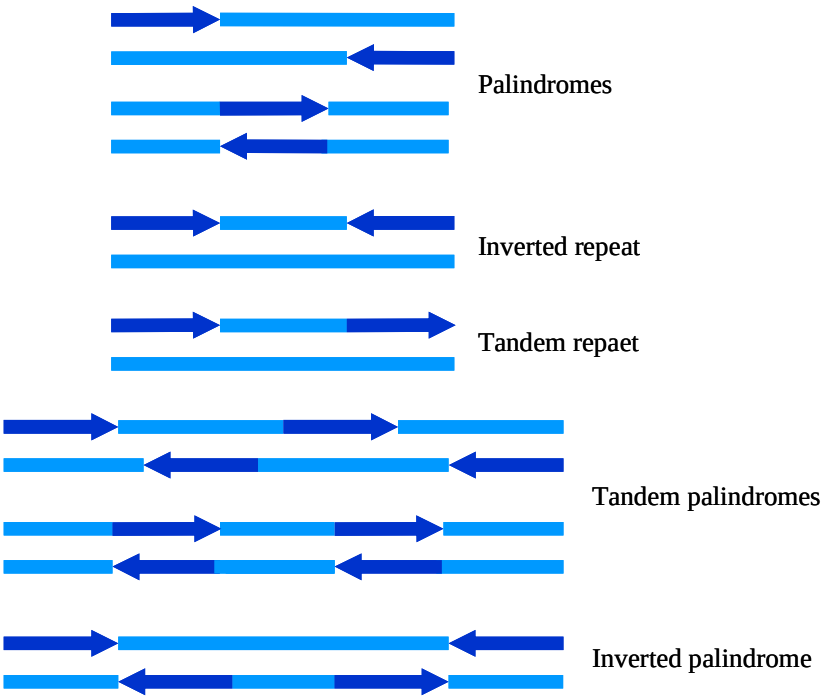

Figure 3: Some possible arrangements of operator sequences. Dark blue arrows indicate the same sequences.

Operators positioned outside of the -60 and +20 positions are termed remote operators (Collado-Vides et al. 1991). A survey of a large number of promoters indicated that remote operators can be found in a number of promoters, for example in the *ara*, *deo*, *gal*, *lac* operons of *E. coli*. These operators, however, have one or more remote or proximal counterparts, and interaction of a regulatory protein with these multiple sites modulates expression of the regulated promoter (Matthews 1992). In contrast, in those operons, where only a single operator is involved in the regulation, the position of the operator is proximal or just slightly remote, as has been shown for promoters regulated by the Fur, LexA, MetJ and PurR proteins (Collado-Vides et al. 1991).

1.1.4. Mechanisms of regulation of gene expression

Without going into too many details, in this section I describe the different type of mechanisms involved in regulation of gene expression.

In promoters where only operators with proximal or slightly remote positions are present the mechanism of gene regulation is quite simple, independent of the type of the DNA-binding motif of the regulatory protein (Figure 4).

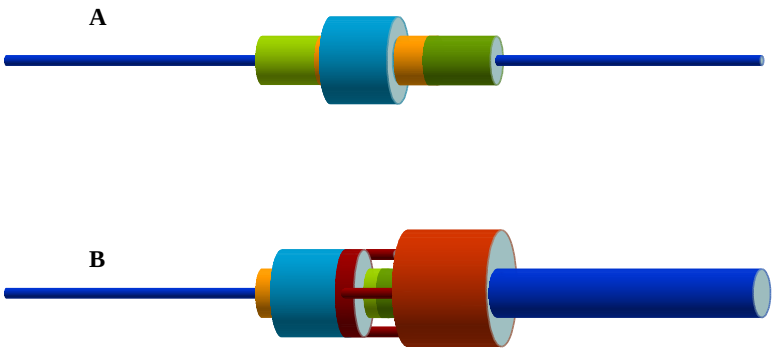

Figure 4: Repression (A) and activation (B) by regulatory proteins with a proximal operator. In repression, the repressor protein (medium blue) binds to the operator (yellow) positioned between the -35 and -10 boxes (light and dark green, respectively) and therefore RNA polymerase cannot initiate transcription. In activation, the activator protein (medium blue) binds to the operator and enhances transcription by RNA polymerase (brown).

In the case of repressors, binding of the protein to the operator site inhibits either formation of the RNA polymerase (RNAP)-promoter complex or the proper functioning of the RNAP, thus preventing expression of the regulated gene (Ishihama 1993; Rojo 1999). On the other hand, activator proteins either contact RNAP or alter promoter conformation to enhance initiation of transcription (Rhodius and Busby 1998).

The mode of regulation is less obvious in promoters in which the operator has a remote, especially extreme, position, and therefore no direct effect on the RNA polymerase is possible to regulate transcription.

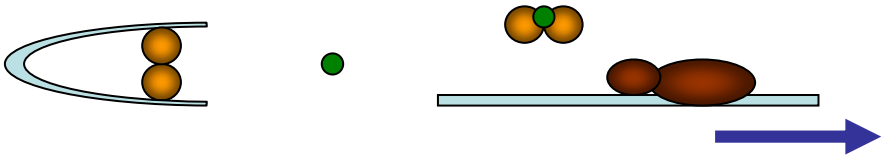

Figure 5: Loop formation in the DNA by a repressor protein bound to a proximal and a remote operator represses gene expression. Upon presence of an inducer molecule, the repressor is released from the operators and transcription is initiated by RNA polymerase. Labelling is the same as in Figure 1.

To date, three basic mechanistic models have been presented describing the mode of regulation when remote operators are involved. In the first model (Figure 5), a regulatory protein binds cooperatively to both remote and proximal operators causing loop formation and thus repression of the regulated promoter (Schleif 1992; Ptashne 2005; Semsey et al. 2005). Under activating conditions, the loop resolves and transcription occurs by RNA polymerase. This type of regulation is known in many bacterial systems; examples include the *ara*, *deo*, *gal* and *lac* operons of *E. coli* (Matthews 1992).

In the second model (Figure 6), a regulatory protein binds to a remote operator and interacts with the RNA polymerase bound to the distant promoter, which interaction causes loop formation in the intervening DNA, and results in activation of the promoter (Gralla 1991; Matthews 1992). For example the NtrC and XylR proteins appear to activate transcription by this mechanism (Su et al. 1990; Pérez-Martin and De Lorenzo 1995). In both models, other proteins such as DNA-specific architectural proteins may promote loop formation (Xu and Hoover 2001).

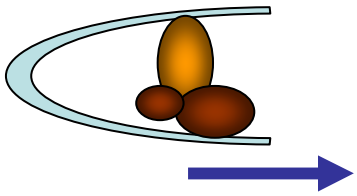

Figure 6: Activation of gene expression occurs by loop formation between an activator protein bound to a remote operator site and RNA polymerase bound to the promoter region. Labelling is the same as in Figure 1.

In the third model described, it is visualised that a regulatory protein bound to a remote operator mediates expression of the distant promoter by modulating the conformation of the DNA between the operator and the promoter (Adhya and Garges 1990). However attractive, there were no examples described in the literature for this model.

1.2 The biology of *Agrobacterium*-plant interaction

1.2.1. *Agrobacterium tumefaciens* is a plant pathogen bacterium

A. tumefaciens is a member of the *Rhizobiacea* family. Virulent strains of the species, which harbour a large, tumour-inducible (Ti) plasmid (Van Larebeke et al. 1974; Figure 7), induce the formation of tumours, otherwise known as crown galls, on a wide range of dicotyledonous plants (Figure 8).

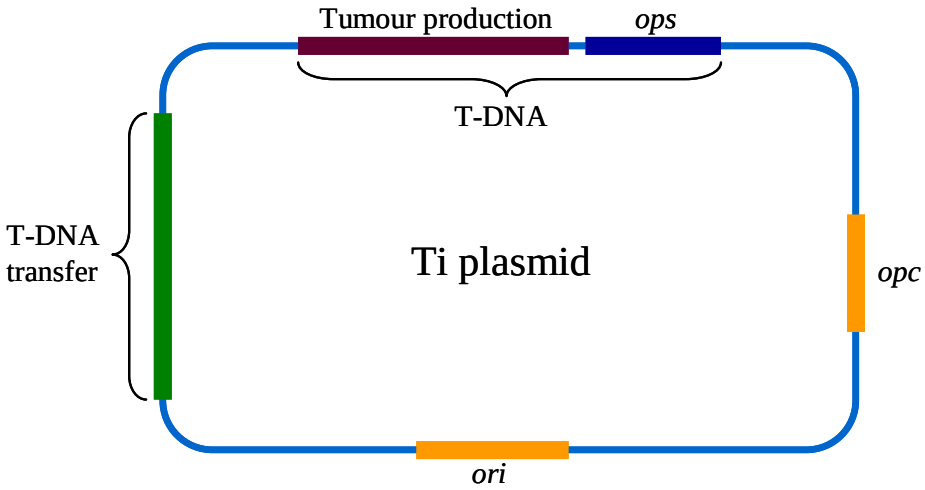

Figure 7. Schematic structure of a generalised Ti plasmid: *ops*, opine synthesis; *opc*, opine catabolism; *ori*, origin of replication.

Figure 8. Crown gall formation on tobacco stem by a virulent strain of *A. tumefaciens*. (The author's experiment.)

A. tumefaciens lives in the rhizosphere of plants where special phenolic compounds secreted by wounded plant tissues serve as chemo-attractants for the bacteria (Brencic et al. 2005). After colonising the wound site, the bacterial virulence (*vir*) genes located in the T-DNA transfer region of the Ti plasmid are induced by the same plant phenols. The products of these

genes are involved in the transfer of a discrete portion of the Ti plasmid (the T-DNA) to the plant cell nuclei where the T-DNA is integrated into the genome (Zupan et al. 2000; Gelvin 2003; Tzfira et al. 2004). Several genes of the integrated T-DNA are involved in formation of tumours on the infected plant. These genes direct the overproduction of the plant growth hormones auxin and zeatin and also increase the plant cells' sensitivity to these phytohormons (Zhu et al. 2000), thus leading to uncontrolled growth of tumours, which are able to proliferate *in vitro* without the addition of hormones. Since *Agrobacterium* is able to transfer its T-DNA into numerous plants (Gelvin 2003), this character has become the basis of one of the most important techniques in plant molecular biology, leading to the *in vitro* transformation of a number of plant species for research and commercial purposes (Valentine 2003).

1.2.2. Opines in *Agrobacterium* biology

Beside genes which encode proteins involved in tumour formation, other T-DNA genes direct the synthesis of unusual low molecular weight molecules, collectively called opines (Dessaux et al. 1993). Although opines can be found elsewhere in nature, their main incidence is in plant tumours induced by *Agrobacteria*. Diverse types of opines can be isolated from naturally occurring tumours found on several plant species (Moore et al. 1997). So far, more than 20 different opines have been described, which can be divided into distinct groups based on their chemical composition. Usually they are conjugates of organic acids and amino acids, sugars and amino acids, or different sugars. Opines are excreted from the tumour cells (Zhu et al. 2000) and can be utilised as carbon and sometimes as nitrogen source by the colonising bacteria (Brencic and Winans 2005). Opines are nutritionally specific, i.e. opine catabolism of the inciting *Agrobacterium* strain is specific to the opine or opines produced by the incited tumour, and this fact evolved in the "opine concept" long time ago (Tempé and Petit 1982). In addition to their role as nutrients, opines have other functions in *Agrobacterium* biology. First, opines serve as chemoattractants for the tumour-inducing *Agrobacterium* strains (Kim and Farrand 1998). Second, certain so called conjugal opines were demonstrated to increase Ti plasmid copy number and to enhance their conjugal transfer (Oger and Farrand 2002; Pappas and Winans 2003; Brencic and Winans 2005). Third, some opines stimulate expression of the *vir* genes (Veluthambi et al. 1989). Whether these functions have a "rank" in *Agrobacterium* biology or just simply co-exist is not known. Opines were also thought to have an important role in both *Agrobacterium* and rhizosphere ecology. Their role in Ti-plasmid conjugative transfer indicates that they might enhance spreading of Ti-plasmids from virulent to

non-virulent bacteria in nature. One part of the original “opine concept”, however, stating that Agrobacteria benefit from opine synthesis, became somehow controversial. Opines were proven to provide a selective advance for utilising bacteria over other rhizosphere microbes (Guyon et al. 1993; Oger et al. 1997; Savka and Farrand 1997). On the other hand it was also shown that opines can be utilised by other rhizosphere microorganisms (Tremblay at al. 1987; Beauchamp et al. 1990; Bergeron et al. 1990; Nautiyal and Dion 1990). Moreover, *Agrobacterium* can initiate T-DNA transfer and opine production without inciting tumour formation (Brencic et al. 2005). It was also shown that certain opines inhibit the growth of Agrobacteria transiently, although it is not known that such sensitive strains exist in nature (Kim et al. 2001). I note here that a more general substance produced by a tumour, indolacetic acid, also can inhibit the growth of Agrobacteria and many other plant-associated bacteria (Liu and Nester 2006). Whether this inhibition by any tumour-produced substance is a defence mechanism of the plant against the invading bacteria is not clear. In this respect, the suggestion that opine catabolism is only a preventive measure by the bacteria against the plant’s defence (Kim et al. 2001) is a bit speculative theory.

A. tumefaciens genes required for opine uptake and catabolism are located on the Ti plasmid, but they are not part of the T-DNA (therefore not transferred into the plant genome), and are specifically induced by the cognate substrate synthesised in the tumour (Bomhoff et al. 1976; Klapwijk et al. 1977; Valdivia et al. 1991). It is worth to note that, although opine synthesis and catabolism are very similar biochemical reactions, some enzymes catalysing these reaction are not related (Zanker et al. 1994) while other biosynthetic and catabolic genes are homologous and very likely have a common evolutionary origin (Kim and Farrand 1996).

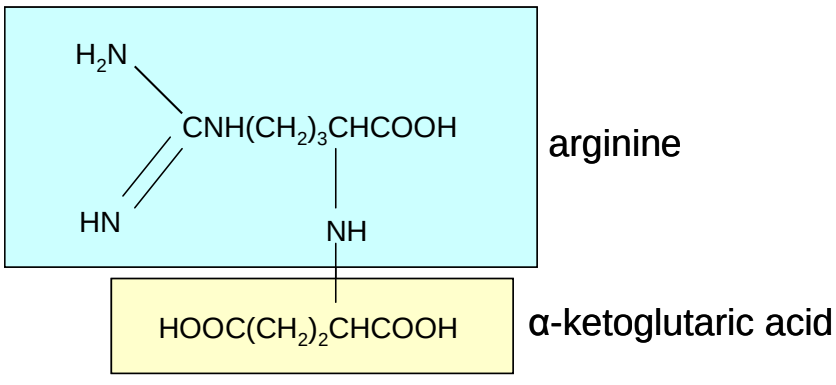

Figure 9. The structure of nopaline.

Nopaline (Figure 9) is one such opine which is synthesised in tumours by the product of the nopaline synthase (*nos*) gene (Depicker et al. 1982), and utilised by *A. tumefaciens* strains containing nopaline-type Ti plasmids, such as pTiC58 or pTiT37 (Holsters et al. 1980; Schardl and Kado 1983; Krishnan et al. 1991).

Prior knowledge about the molecular biology of nopaline uptake and catabolism was very limited before I started my work. Genetic and functional analysis of the nopaline catabolism (*noc*) region of the Ti plasmids pTiC58 and pTiT37 indicated that several genes are involved in nopaline uptake and catabolism (Holsters et al. 1980). Experiments in our laboratory and elsewhere suggested that the nopaline-inducible genes of these plasmids are divided into two operons separated by a portion of DNA of unknown function and that these inducible genes might be regulated by a single gene expressed constitutively and transcribed divergently from the regulated genes (von Linting et al. personal communication; White et al., unpublished results).

Based on these preliminary results, the aims of my work were: (i) to study the structure of the putative regulator gene of nopaline catabolism of the Ti plasmid pTiT37; (ii) to learn the function of the regulatory protein; and (iii) to study regulatory elements and mechanisms involved in the regulation of nopaline catabolism in *A. tumefaciens*.

2. Materials and methods

Particular strains and plasmids, and growth of strains are described in the following publications: Marincs and White 1993; Marincs and White 1994; Marincs and White 1995; Marincs and White 1996.

Molecular techniques, such as isolation of DNA, restriction digestion, DNA electrophoresis, cloning, sequencing and transformation were done by standard methodologies (Sambrook et al. 1989). DNA topoisomer analysis was performed as described (Marincs and White 1995).

Expression, isolation and gel electrophoresis of proteins and measurement of enzyme activity are described in details in the following publications: Marincs and White 1993; Marincs and White 1994; Marincs and White 1995.

Protein-DNA interactions were studied by gel-retardation and DNaseI footprinting (Marincs and White 1993).

Introduction of mutations into DNAs were done by site-directed mutagenesis (Marincs and White 1994).

3. Results and discussion

3.1. Characterisation of the *nocR-nocP* region

Genetic and functional analysis of the nopaline catabolism (*noc*) region of *A. tumefaciens* Ti plasmids indicated that several genes are involved in nopaline uptake and catabolism (Holsters et al. 1980). It was postulated that the *noc* genes of plasmids pTiC58 and pTiT37 are regulated by a single gene. Previously, several genes had been localised in the *noc* region of pTiT37 using a *lac* fusion transposon (Stachel et al. 1985) and it was established that two of these genes, *nocR* and *nocP* (formerly also called *nocB*) are divergently transcribed and that *nocR* is involved in the nopaline-induced expression of *nocP* (White et al. unpublished results).

To characterise the *nocR-nocP* region of pTiT37, a 2.3 kbp PstI fragment containing the region was sequenced. A 900 bp open reading frame (ORF) and a putative promoter and a ribosome binding site (RBS) for this ORF were identified by analysing the sequence (Figure 10 and Appendix 1: Figure 1). It was predicted that the identified ORF corresponds to the *nocR* gene. Comparison of the deduced protein sequence with protein databases revealed that the putative NocR protein is related to the members of the LysR family of prokaryotic transcriptional activators (Henikoff et al. 1988). The LysR family is thought to be the largest family of prokaryotic DNA-binding proteins. Amongst 314 DNA-binding proteins of *E. coli* alone, there are 18 verified and 27 predicted members of the LysR family (Zaim and Kierzek 2003), but LysR-type proteins are exist in diverse bacterial genera (including *Rhizobiaceae*), Archaea and even algal chloroplast (Schell 1993). The highest sequence homology between LysR-type proteins can be found at their N-terminal domains where the HTH DNA-binding motives are located (Henikoff et al. 1988). LsyR and NocR share an overall 32% homology (Marincs and White 1993), while the *nocR* gene of pTiT37 is identical to of that of pTiC58 (von Linting et al. 1991). The predicted NocR protein of pTiT37 has a HTH DNA-binding motif (Harrison 1991) near the N-terminal end (Figure 10 and Appendix 1: Figure 1).

The *nocR* gene was cloned into the *tac* promoter expression vector pKK223-3. Using denaturing protein gel electrophoresis, expression of a 31.5 kDa protein was revealed (Appendix 1: Figure 2), and this molecular mass is in very good correlation with the 31.2 kDa molecular mass calculated from the deduced amino acid sequence.

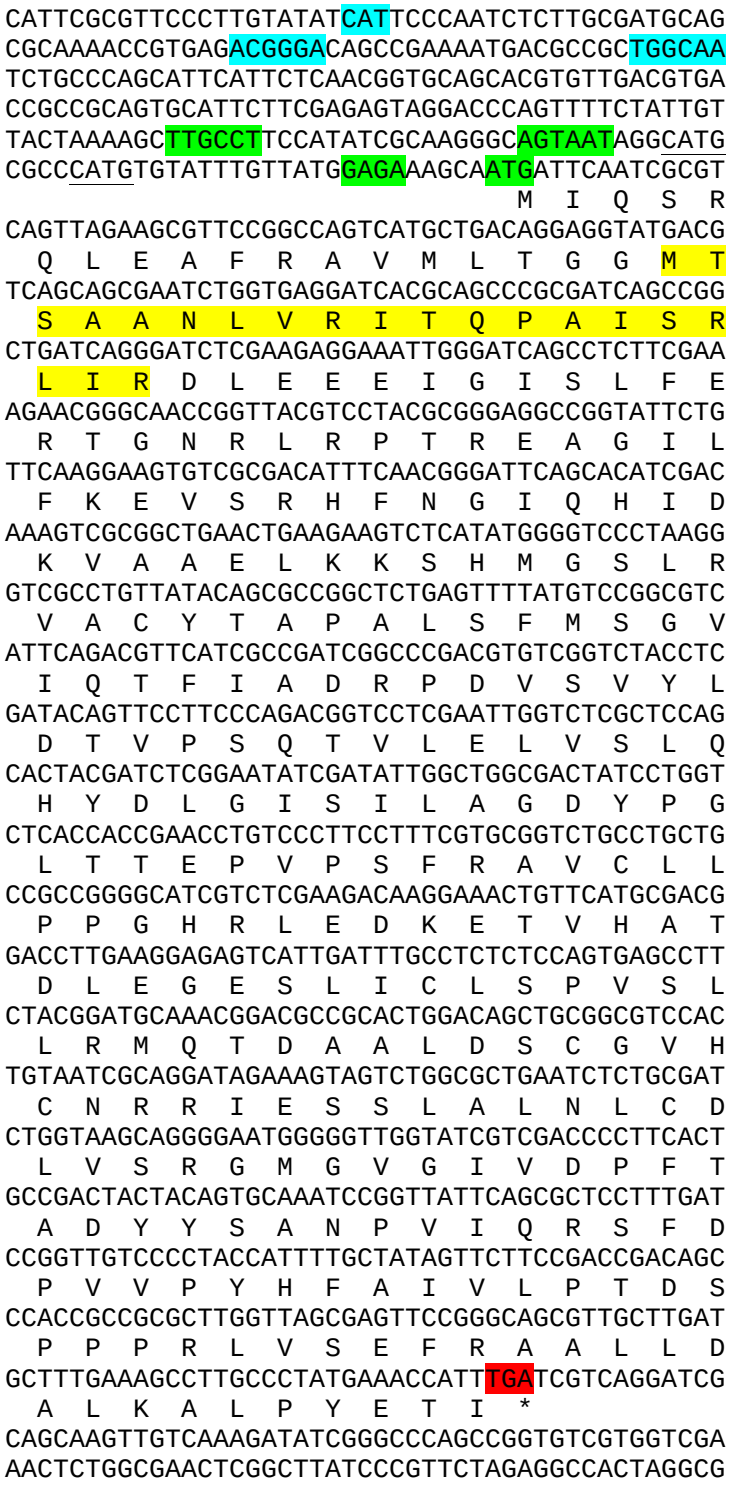

Figure 10. Nucleotide and amino acid sequence of the *nocR* gene and its 5' and 3' regions. Start and stop elements of the *nocR* gene are highlighted in green and red, respectively. Start elements of the divergently transcribed *nocP* gene are highlighted in blue. Repeats of the *noc* operator are underlined. In the deduced amino acid sequence of NocR, the putative helix-turn-helix motif is highlighted in yellow. (Adapted from Marincs and White 1993)

In addition, the first ten amino acids of the expressed protein, determined by micro-sequencing, were identical to the predicted amino acids (Marincs and White 1993).
Database search with another, partial ORF that was found in the sequenced region and corresponds to the regulated *nocP* gene indicated that it encodes a putative ABC-transporter protein, which is very likely involved in the uptake of nopaline (Marincs and White, unpublished result).

3.2. Regulatory functions of the NocR protein

Sequence analysis revealed that the *nocR* and *nocP* genes of the *noc* operon of plasmid pTIT37 are divergently transcribed (Marincs and White 1993). Similar gene arrangements have been found in a large number of prokaryotic operons (Beck and Warren 1988), where one of the divergently transcribed genes regulates expression of the other gene(s). In addition, many members of the LysR family are known to auto-regulate their own synthesis (Schell 1993). Since the NocR protein displays homology with members of the LysR family and has a putative DNA-binding HTH motif (see above), it was obvious to investigate whether it regulates expression of *nocP* and auto-regulates its own synthesis. To study these aspects, a number of clones of the *nocR-nocP* region were constructed, in which the coding regions of either *nocP* or *nocR* or both were replaced with reporter genes. The firefly luciferase (*luc*, Greer and Szalay 2002) and the *E. coli* β-glucoronidase (*gusA*, Jefferson et al. 1987) genes were used to replace the *nocP* and *nocR* coding sequences, respectively (Appendix 2: Figure 1B). Hence, transcriptional activity of the *nocP* and *nocR* promoters could be extrapolated by measuring activities of the reporter enzymes. Using these constructs, expression of both the *nocP* and *nocR* genes was investigated in *A. tumefaciens* strains either harbouring or lacking plasmid pTiT37, which provides the NocR protein in *trans* (Marincs and White 1994).
In the absence of nopaline and the presence of the NocR protein, very low level of luciferase activity was detected indicating transcriptional repression of the *nocP* promoter driving the *luc* reporter gene. In contrast, luciferase activity was about 100-fold higher in the presence of both nopaline and NocR, indicating either de-repression or activation of the *nocP* promoter. The results also indicated, however, that the NocR protein is not needed for the expression of *nocP* since in a clone containing the putative *nocP* promoter fused to the *luc* reporter gene high constitutive luciferase activity was detected (Table 1 and Appendix 2: Table 1). It was also revealed that NocR auto-represses its own synthesis since activity of the GusA protein was 9 to 39-fold higher (depending on the substrate used for measurements) in the absence

than in the presence of NocR. This auto-repression function of NocR was independent of nopaline (Table 1 and Appendix 2: Table 2).

Table 1. Summary of the expression data for the *nocR* and *nocP* genes

	Plasmid genotype	NocR protein[a]	*nocR*[b] -nop[d]	*nocR*[b] +nop[d]	*nocP*[b] -nop[d]	*nocP*[b] +nop[d]
1.	LPpRpOG	+	1	1	1	128
2.	LPpRpG	+	1	1	15	12
3.	LPpRpOG	-	9	9	10	11
4.	LPpRpG	-	1	1	13	12
5.	LPpRpO	+	-	-	47	140
6.	LPpRpO	-	-	-	47	54
7.	LPpRp	+	-	-	162	172
8.	LPpRp	-	-	-	167	175
9.	LPp	+	-	-	40	39
10.	LPp	-	-	-	36	38
11.	RpOG	+	1	1	-	-
12.	RpOG	-	35	39	-	-

In the plasmid genotype column: L = *luc*; Pp = *nocPp* (promoter), Rp = *nocRp* (promoter); O = *nocPo* (operator); G = *gusA*.

[a] = the presence and absence of the NocR protein in *trans* are labelled + and -, respectively.

[b] = normalised expression of the *gusA* and *luc* genes replacing the *nocR* and *nocP* coding regions.

[d] = the absence and presence of nopaline are labelled by –nop and +nop, respectively.

In the experiments shown in lanes 11 and 12, a different substrate was used to measure β-galactosidase activity causing a different relative expression level in lane 12 compared to lane 3 (Adapted from Marincs and White 1994, 1995)

From these results it was concluded that NocR is a diffusible *trans*-acting factor with two regulatory functions. Firstly, NocR negatively auto-regulates its own synthesis, which is quite a common character amongst LysR-type proteins. For example, AsnC, CysB, NahR, OccR and OxyR regulate the expression of their own genes in this way (Kölling and Lother 1985; Schell and Faris 1987; Christman et al. 1989; Habeeb et al. 1991; Kredich 1992). Secondly, NocR both represses and activates expression of *nocP*.

In the absence of a repressor gene the regulated promoter is expressed constitutively (Lewin 1990), and it was shown that in the absence of the *nocR* gene *nocP* is constitutively expressed (Marincs and White 1994); thus NocR complies with the criteria that defines a repressor. On the other hand, induction of *nocP* in the presence of both NocR and nopaline (Marincs and White 1994) indicates the activator-like character of NocR. Although 22% of all regulatory proteins are dual regulators in *E. coli* (Pérez-Rueda and Collado-Vides 2000), such a character is very rare amongst LsyR-type proteins whose functions have been confirmed based on the criteria for repressors and activators (Lewin 1990). The TfdS protein has been shown to regulate expression of *tfdB* both negatively and positively and, similarly to NocR, TfdS needs a co-inducer molecule for the activator function (Kaphammer and Olsen 1990). However, the level of repression by TfdS is only about 40% of the full expression, in contrast to the 98% repression observed with NocR.

3.3. NocR-DNA interaction in the *noc* operon

Sequencing of the *nocR-nocB* region of plasmid pTiT37 revealed that the two genes are divergently transcribed and are separated by a 231 bp region (Marincs and White 1993). *In silico* analysis of the intervening sequence revealed the position of putative promoters for both *nocR* and *nocP*, and functional analysis confirmed the position of the *nocP* promoter (Marincs and White 1993, 1994). Since NocR was shown to regulate expression of both *nocP* and *nocR* and it has a HTH DNA-binding motif it was obvious to conclude that the NocR protein might interact with the *nocR-nocP* promoter region in order to facilitate its regulatory functions.

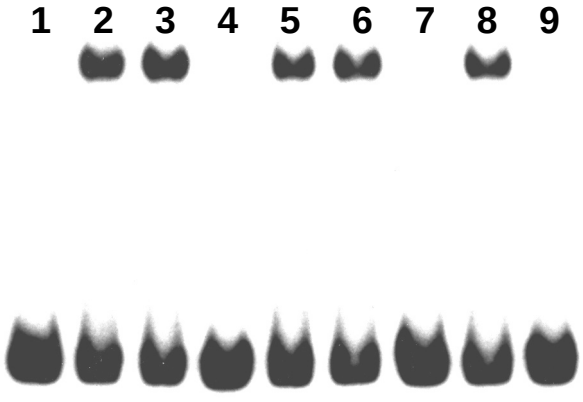

Figure 11. Competition analysis of the interaction between the NocR protein and the *noc* promoter region. Only those fragments of the region compete with the binding which contain one particular 76 bp sub-fragment (lanes 4, 7 and 9). (From Marincs and White 1993)

I have investigated this possibility using gel retardation and DNaseI footprinting techniques (Marincs and White 1993).

These experiments revealed that NocR binds exclusively to a 76 bp sub-fragment of the 261 bp long *nocR-nocP* promoter region (Figure 11 and Appendix 1: Figures 4, 5 and 8A). This sub-fragment covers the sequence between the 5' end of the -35 hexamer of the *nocR* promoter and the ATG start codon of the *nocR* gene (Marincs and White 1993).

Nopaline had no effect on the formation of the NocR-DNA complex since NocR binds both in the absence and in the presence of nopaline to its target DNA. Nopaline, however, affected the migration of the protein-DNA complex; increasing concentration of nopaline resulted in increased migration of the NocR bound DNA (Appendix 1: Figure 6). Octopine, a related opine, and precursors of nopaline had only tiny or no effect on migration of the complex, respectively (Appendix 1: Figure 7).

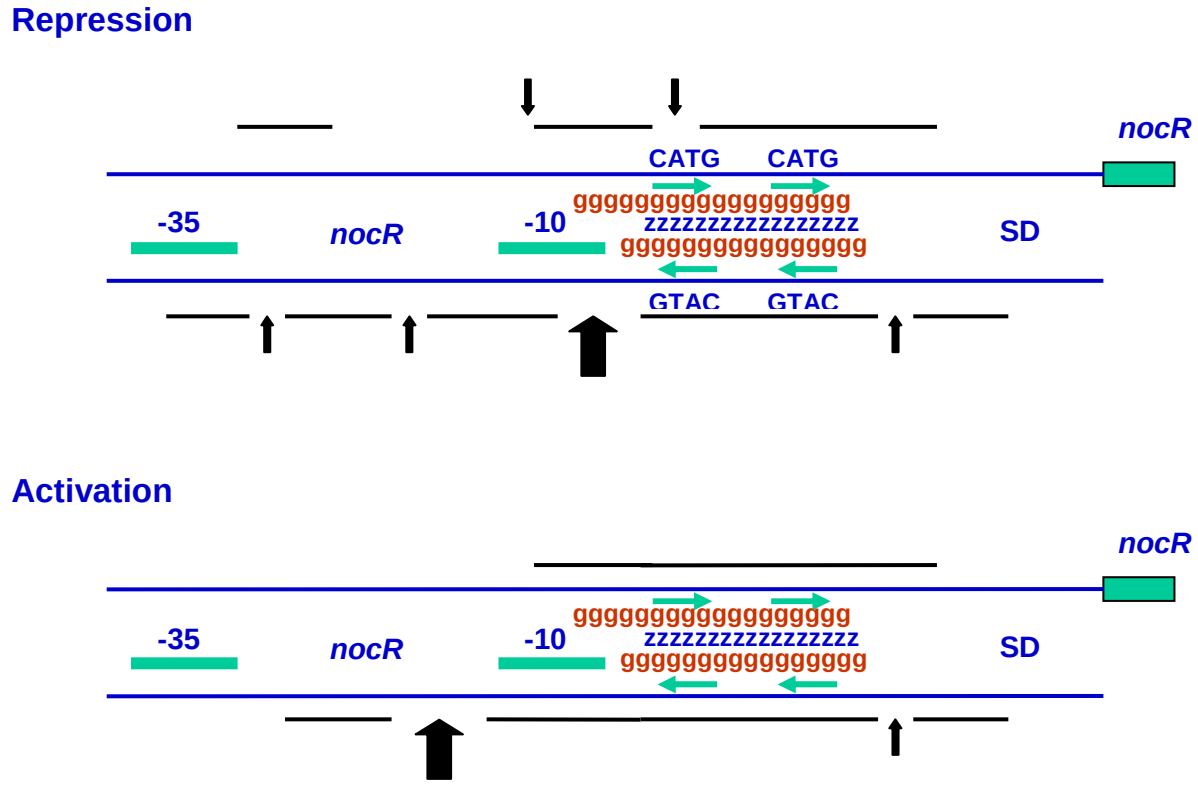

Figure 12. DNaseI footprinting of the *nocR* promoter region in the absence (repression) and presence (activation) of nopaline and characteristic sequences in the fragment. Black lines and arrows label the NocR-protected sequences against DNaseI and hypersensitive sites, respectively. Green arrows label the CATG palindromes of the *noc* operator; small case g and z letters label the gyrase consensus and the alternating purine-pyrimidine sequences, respectively. (Adapted from Marincs and White 1993)

DNaseI footprinting also indicated that the NocR protein protects the *nocR* promoter but no other sequences in the 261 bp *noc* promoter region and that nopaline causes characteristic changes in the DNaseI hypersensitivity pattern of the protected fragment (Figure 12 and Appendix 1: Figure 8). Closer examination of the *nocR* promoter region to which the NocR protein binds revealed the presence of a twelve base-pairs long putative operator with the sequence of $CATGN_4CATG$ (Figure 10). The first basepair of this highly symmetric tandem palindrome is positioned 3 bp downstream of the -10 region of the *nocR* promoter. Deletion of this sequence from the *nocR* promoter resulted in no binding of NocR to its target DNA region *in vitro* (Appendix 2: Figure 2), indicating the importance of this particular sequence in binding of NocR to the *noc* promoter region.

In comparison with NocR, the interaction of closely related proteins with their target DNA shows both differences and similarities. For example, while nopaline influenced the migration of the DNA-protein complex in pTiT37 (Marincs and White 1993), it had no effect in another nopaline-type plasmid, pTiC58 (von Lintig et al. 1994). Octopine, a related opine, affected migration of the DNA-protein complex in the octopine-type plasmids, pTiB6S3 and pTiA6 (Wang et al. 1992; von Lintig et al.1995) and even in the heterelogous nopaline-type plasmid, pTiC58 (von Lintig et al. 1995), but its effect was marginal in pTiT37 (Marincs and White 1993). The footprint of NocR of pTiT37 was also different compared to of that of the other plasmids mentioned (Wang et al. 1992; von Lintig et al. 1994).

3.4. Specific structural elements of the *noc* operon

There are two structural features in the *noc* operon of pTiT37, which make it quite unique compared to other opine catabolism and LysR-type proteins-regulated operons.

First, in contrast to other opine catabolism operons, where the promoters of the regulatory and regulated genes overlap (Wang et al. 1992; Von Lintig et al. 1994), in the *noc* operon of pTiT37 there is a relatively large distance between the regulated *nocP* promoter and the binding site of the NocR protein (Figure 10). The -10 and -35 regions of all known opine catabolism promoters are highly homologous to the *E. coli* σ^{70} -10 and -35 hexamers (Lisser and Margalit 1993), thus it was predicted that the *nocP* promoter should also display such a homology. Despite this, no such homologies were found overlapping the *nocR* promoter of pTiT37 (Marincs and White, 1994). *In silico* and functional analyses, however, indicated the presence of a promoter for *nocP* 131 bps upstream of the *noc* operator. This promoter is highly homologous to the *E. coli* σ^{70} consensus promoter in the critical first and sixth

positions, although the distance between the -10 and -35 hexamers is 19 bp instead of the optimal 17 (Figure 10).

Second, the highly symmetrical $CATGN_4CATG$ 12 bp tandem palindrome sequence (Figure 10), which was proven to serve as an operator for regulating expression of the *nocP* gene, is quite unique amongst operons regulated by LysR-type proteins. A single $TN_{11}A$ sequence was predicted to function as an operator for genes regulated by members of the LysR family (Goethals et al. 1992), including other opine catabolism operons (Von Lintig et al. 1994). Similar sequences were also located in the *noc* operon of pTiT37, but because of their relatively large number, it was concluded that this type of sequence is not an operator in pTiT37, in particular that the operator function of the $CATGN_4CATG$ tandem palindrome was proven both *in vitro* and *in vivo*.

3.5. Molecular interactions in the *nocR-nocP* region

In vitro and *in vivo* studies described above revealed the function of the NocR protein and its interaction with its target sequence, but the mechanism by which NocR regulates gene expression in the *noc* region still needed an explanation.

To study this mechanism, I have investigated the effect of the putative operator sequence on expression of the *nocR* and *nocP* genes. A series of plasmids with the reporter genes *luc* and *gusA* that replaced the coding portion of *nocP* and *nocR* genes, respectively, were constructed. From some of these constructs the coding sequence driven by the *nocR* promoter and/or the putative operator were also deleted (Marincs and White 1995). These plasmids were then introduced into *A. tumefaciens* strains having or lacking the *nocR* gene. From these experiments it became clear once again that nopaline does not affect expression of the *nocR* gene and its expression is auto-regulated by the NocR protein (Table 1).

The picture in the case of *nocP* was more complicated. Summarising expression of the *nocP* gene (Table 1 and Appendix 3: Table 1), four distinct levels were observed depending on the presence/absence of NocR (a *trans* factor) and the operator and divergent transcription (*cis* factors):

1. Full repression in the presence of all *cis* and *trans* factors, which is fully de-repressed in the presence of nopaline.

2. Low level expression in the absence of either NocR or the operator or both, which is not influenced by nopaline.

3. Medium level expression in the absence of divergent transcription, which is fully de-repressed in the presence of nopaline and NocR.

4. Full de-repression in the absence of both the operator and divergent transcription either in the presence or absence of NocR.

From these results two major conclusions could be drawn. First, the results indicated that transcription from the divergent *nocR* gene itself can repress expression of *nocP* to a certain extent, even in the absence of NocR. One possible explanation for this phenomenon is that closely spaced promoters such as *nocRp* and *nocPp* compete for the RNA polymerase and that competition influences their expression (Goodrich and McClure 1991). Expression of *nocP*, however, was about the same in those strains where expression of *nocR* was different, making the competition theory quite unlikely. Another possibility is that transcription from the *nocR* promoter influences expression from the divergently transcribed *nocP* promoter. According to the twin-domain model, transcription generates negative supercoils behind the transcription complex in a topologically closed domain (Liu and Wang 1987; Figueroa and Bossi 1988; Tsao et al. 1989; Deng et al. 2005). It was suspected that the generated torsional stress may affect functions of nearby sequences (Tsao et al. 1989; Wang and Giaever 1988; Travers 1989, Deng et al. 2005). In addition, it is known that the supercoil-generating effect of transcription depends on the transcript's length (Brill and Sternglanz 1988). In the absence of the coding sequence downstream of the *nocR* promoter, expression of *nocP* was elevated by about 5-fold independent of the presence of the NocR protein (Table 1 and Appendix 3: Table 1), indicating that divergent transcription has a reducing effect on the expression of *nocP*.

The second conclusion was that the $CATGN_4CATG$ sequence is the operator for the NocR protein. The results, however, also indicated that the operator has different effects on the expression of *nocP*, depending on other factors. In the absence of the NocR protein, expression level of *nocP*, although quite low, is the same in either the presence or absence of the operator when the divergent transcription of *nocR* is intact. In contrast, when transcription of *nocR* is abortive, expression of *nocP* is much higher in the absence then in the presence of the operator sequence. The question is that why the operator sequence has such an interesting effect.

There are two features which might cause the observed effects of the operator. First, *in silico* analysis of the operator and its immediate neighbourhood indicated the presence of an 18 bp long alternating purine-pyrimidine sequence with one bp out of alternation (Figure 12 and Marincs and White 1993). Such sequences are able to form Z-DNA under various conditions (Herbert and Rich 1999). Alternating purine-pyrimidine, putative Z-DNA forming sequences are relatively common in eukaryotes (Herbert and Rich 1999). Although Z-DNA studies were almost exclusively performed using artificial sequences in prokaryotes (Herbert and Rich

1999), natural sequences capable of forming Z-DNA are also present in bacteria (Ansari et al. 1992). It is possible, that such a sequence overlapping the *noc* operator might exist in both B and Z-forms. An indication that the *noc* operator or its close vicinity might form Z-DNA is that DNaseI footprinting experiments revealed that the alternating purine-pyrimidine sequence in the *noc* operator is bordered one particularly strong hypersensitive site, and this site migrates about ten base-pairs upstream under inductive conditions, indicating the expansion of a possible Z-DNA stretch (Figure 12 and Appendix 1: Figure 8). A recent study has shown that at B-Z junctions a single base-pair is broken and flipped out of the double helix (Ha et al. 2005) which can result in such nuclease hypersensitivity that was found in the *noc* operator.

The second feature is a sequence highly homologous with the *E. coli* gyrase recognition site (Lockshon and Morris 1985) overlapping the *noc* operator (Figure 12 and Marincs and White 1993). Gyrase is known to introduce and also remove supercoils to or from DNA (Champoux 2001). When the *noc* operator sequence was cloned into a high copy number plasmid (Marincs and White 1993, 1994), the topoisomer pattern (i.e. the distribution of plasmid molecules with different supercoil level) of the plasmid became altered if activity of gyrase was inhibited (Figure 13 and Appendix 3: Figure 3).

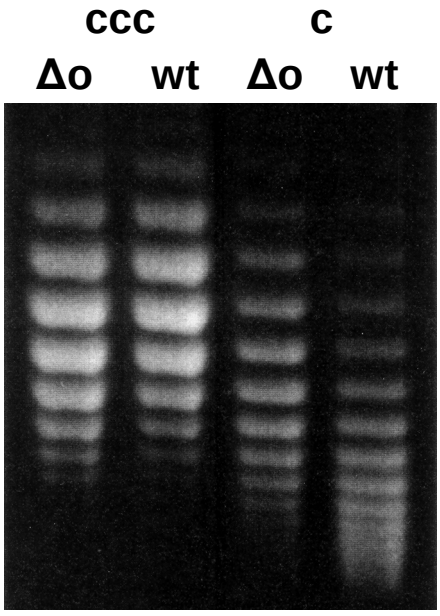

Figure 13. Topoisomer pattern of plasmids lacking (Δo) and carrying (wt) the noc operator under less (ccc) and more (c) relaxed conditions. (From Marincs and White 1995)

Different explanations can be behind the above described observations. For example, as mentioned earlier, supercoils generated by the transcription of the *nocR* gene can result in B-Z

transition of the *noc* operator, since it is well know that B-or Z-DNA can exist under different conditions *in vivo*, for example depending on the level of supercoiling (Rich and Zhang 2003). And vice versa, transition of a DNA sequence from B- to Z-form can remove supercoils (Rich and Zhang 2003).

Altogether, the different levels of expression of *nocP* depending on the absence or presence of the different structural *cis* elements might indicate the involvement of DNA supercoiling, Z-DNA formation and complex molecular interactions in the regulation of the *nocP* gene.

3.6. Local structures involved in the regulation of the *noc* operon

The first local property, which might be involved in the expression of the *nocP* gene is DNA supercoiling. Alteration in superhelicity is a possible mechanism that was proposed to regulate gene expression at a distance a long time ago (Smith 1981; Pruss and Drlica 1989). It is well know that expression of a number of bacterial genes is sensitive to the level of supercoiling, and it was shown that about 7% of the *E. coli* transcriptome is affected by changes in the level of supercoiling (Peter et al. 2004). Negative supercoiling can be perturbed by environmental factors, mutations in genes encoding DNA topoisomerases and certain antibiotics, and it was shown to modulate transcription from particular promoters (Deng et al. 2005). It was thought therefore, that it affects gene expression in a global manner. However, as suggested recently, RNA polymerase senses the local superhelical parameters and not the global supercoiling level (Travers and Muskhelishvili 2005), thus transcription is more likely influenced by local and not global superhelicity.

In this respect, those pieces of evidences which indicate that local supercoiling might have a role in expression of the *nocP* gene are quite significant and listed below.

1. The *nocP* promoter has a 19 bp long spacer between its -10 and -35 hexamers, and the -10 hexamer is GC rich (Marincs and White 1994), which characters have been reported for supercoiling-sensitive promoters (Rosenberg and Court 1979; Condee and Summers 1992).
2. The *noc* operator overlaps a putative gyrase recognition site (Marincs and White 1993).
3. The *noc* operator influences plasmid supercoiling *in vivo* under certain conditions and has a repressing effect on the expression of *nocP* (Marincs and White 1995).
4. Divergent transcription, known to generate negative supercoils behind the transcription complex, represses the *nocP* gene (Marincs and White 1995).
5. Carbon starvation which relaxes DNA (Balke and Gralla 1987) and high osmolarity which increases supercoiling (Hsieh et al. 1991) enhances and reduces expression of the *nocP* promoter, respectively (Marincs, unpublished results).

The second local feature which might be involved in the expression of the *nocP* gene is Z-DNA and the evidences supporting this possibility are listed below.

1. The *noc* operator overlaps with a putative Z-DNA forming sequence (Marincs and White 1993).

2. The *noc* operator displays strong nuclease hypersensitivity which can be the indicator of B-Z junctions (Marincs and White 1993).

3. Under high salt concentration the *noc* operator DNA displays some changes in its circular dichroism spectrum (Marincs, unpublished result), which is characteristic for B to Z-DNA transition (Herbert and Rich 1999).

3.7. Model for the regulation of gene expression in the *noc* operon

As it is described above in section 3.2, the NocR protein has two regulatory functions. First, it auto-regulates its own synthesis, and second, it regulates expression of the *nocP* gene (Marincs and White 1994).

The possible mechanism of auto-regulation is quite clear. Similar to many LysR-type regulatory proteins, NocR binds to an operator sequence just downstream of the *nocR* promoter under either repressive or inductive conditions, physically preventing the transcription from the *nocR* promoter by RNA polymerase, thus repressing its own synthesis. In the absence of NocR, expression from the *nocR* promoter is de-repressed.

Expression of *nocP* is more intriguing, and it is possible that a number of factors are involved in its regulation and a different mode of regulation is also possible for its repression and activation.

By definition, the position of the *noc* operator can be termed remote, and no other operator was found in either proximal or a closer remote position relative to the *nocP* promoter. Consequently those regulatory models, described in more details in the Introduction section, in which interaction between a regulatory protein and the RNA polymerase is required to regulate gene expression, are not applicable for the *noc* operon, in particular because the NocR protein binds to its operator in both absence and presence of the inducer molecule and there is no evidence for DNA bending in the *nocR-nocP* promoter region.

The relative orientation of the -10 and -35 hexamers of σ^{70} promoters is very important for their interaction with the RNA polymerase (Wang and Syvanen 1992). The *nocP* promoter has a 19 bp spacer between its -10 and -35 hexamers and consequently the spatial conformation of the hexamers is suboptimal for transcription (Figure 14). It is very likely that under repressive condition this suboptimal spatial conformation of the *nocP* promoter is

maintained by two factors: (i) binding of NocR to the *noc* operator 131 bp upstream of the *nocP* promoter, and (ii) transcription of the *nocR* gene, which generates negative supercoils toward the *nocP* promoter.

Under activating conditions, i.e. in the presence of nopaline, expression of the *nocP* promoter was much stronger than in the absence of nopaline. For that enhanced expression, its promoter must be in the correct spatial conformation (Figure 14), and the question is that how this can be achieved.

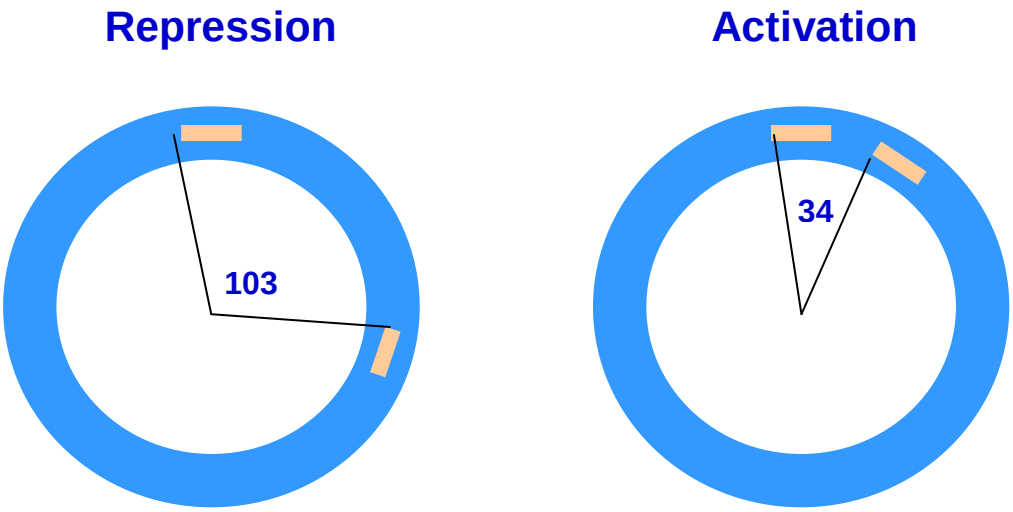

Figure 14. Relative spatial orientation of the -35 and -10 hexamers of the *nocP* promoter under repressive and activating conditions. Numbers indicate the relative angle between the first 5' basepair of the hexamers.

It has been described above, that certain conditions, which can decrease the level of DNA supercoiling, enhance expression of the *nocP* promoter. It is possible therefore, that activation of the *nocP* promoter, i.e. bringing its hexamers into the correct spatial conformation, may occur by a mechanism which decreases the level of local supercoiling. It is known that B- to Z-DNA transition leads to supercoil relaxation (Rich and Zhang 2003), and it was found that the *noc* operator, which overlaps with an 18 bp alternating purine-pyrimidine sequence, was able to reduce superhelicity of a plasmid depending on the overall level of DNA supercoiling, thus it was proposed that the *noc* operator might be to form Z-DNA in vitro (Marincs and White 1995).

To activate the *nocP* promoter, i.e. to bring the -10 and -35 hexamers into the optimal spatial conformation for transcription, a -68.6^{o} rotation between the first residues of the hexamers is

needed (Figure 14). As it was described in more details (Marincs and White 1996), this can be achieved by removing about 1.4 helical turns from the *noc* promoter region. How can these 1.4 helical turns be removed? When certain numbers of base-pairs undergo B- to Z-DNA transition, removal of helical turns occurs (Sinden 1994a). In theory, to remove the 1.4 helical turns, flipping of eight base-pairs from B- to Z-form would be sufficient. Since the *noc* operator overlaps with an 18 bp long putative Z-DNA forming sequence, even partial Z-DNA formation in this sequence would be able to remove the number of helical turns needed to activate the *nocP* promoter (Marincs and White 1996). However, because of their base-composition, natural purine-pyrimidine sequences might not be so efficient in Z-DNA formation as synthetic CG stretches. Thus the entire *noc* operator or even the whole 18 bp purine-pyrimidine sequence might be part of the transition. Another strong possibility is that the *noc* operator is already in Z-DNA form under repressing conditions. This is indicated by the strong DNaseI hypersensitive site at the 5' end of the alternating 18 bp in the absence of nopaline. Under activating conditions, i.e. in the presence of nopaline, this hypersensitive site moves 11 bp toward the *nocP* promoter. This can indicate an expansion of the sequence which is in Z-DNA form when the *nocP* promoter is activated. That expansion also would be sufficient to remove the 1.4 helical turns required for activation of *nocP*. Clearly, nopaline causes some conformational changes in the NocR protein-*noc* operator complex, which changes might indicate either a B- to Z-DNA transition or an expansion of a Z-stretch.

It is not known, however, whether the NocR protein could bind to any DNA which is in Z-conformation. There are a number of proteins in both prokaryotes and eukaryotes which were demonstrated to bind Z-DNA (Rich and Zhang 2003), and therefore it cannot be excluded that NocR has the same character.

It is also not clear whether DNA gyrase has any role in the regulation. There is a sequence very highly homologous to the *E. coli* gyrase recognition site overlapping with the *noc* operator, and theoretically this sequence might have a role in adjusting the local level of supercoiling by interacting with a gyrase.

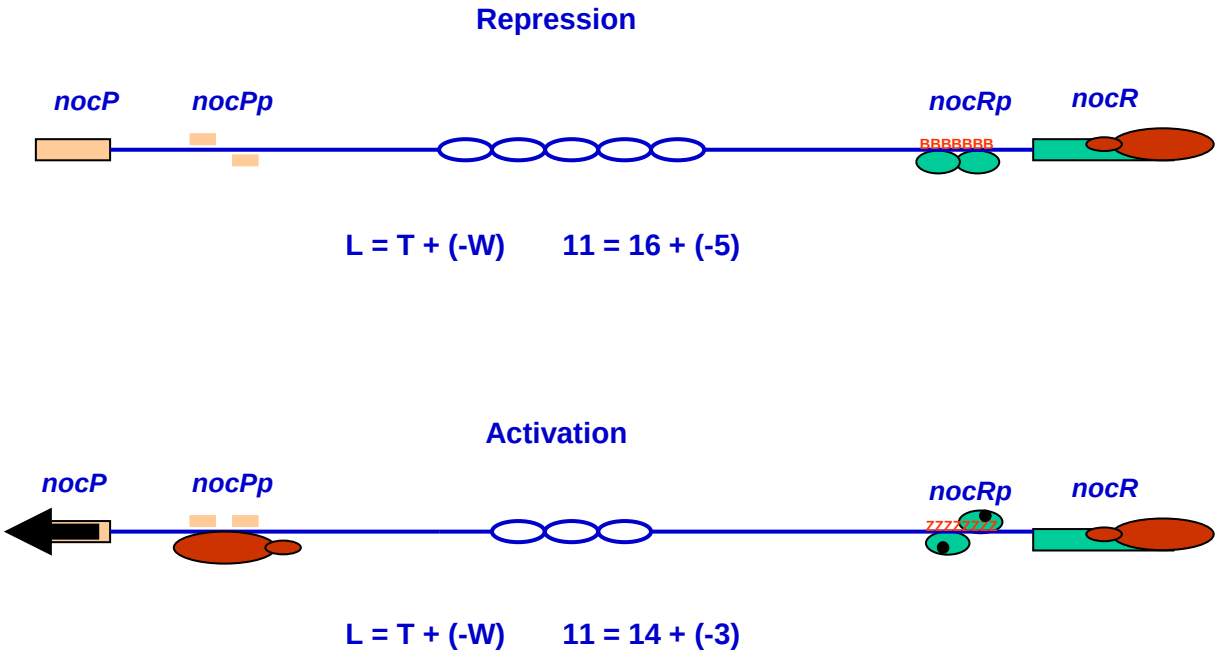

Figure 15. Model for the activation of the *nocP* promoter. The NocR protein and RNA polymerase are labelled by green and brown ellipses. Red B and Z letters label the *noc* operator in B and Z-DNA forms. Open ellipses represent DNA supercoils. Small black circles represent nopaline. In repression, the NocR protein binds to the operator and maintains it in either a B or a shorter Z-DNA conformation. Thus the given number of local negative supercoils in the region fixes the hexamers of the *nocP* promoter in the wrong spatial conformation for transcription. In the presence of nopaline, the NocR protein causes a conformational change of the operator sequence, which results in removal of both supercoils and helical turns in the region bringing the hexamers of the *nocP* promoter into that correct spatial conformation which can be transcribed by the RNA polymerase. (Adapted from Marincs and White 1995)

I have described above that activation of the *nocP* promoter is possible by removing some helical turns in the region by forming Z-DNA upstream of *nocP*. It was also described earlier that the *nocP* promoter is more efficiently transcribed when DNA is more relaxed, i.e. has less supercoils. Furthermore, formation of Z-DNA is known to remove not only helical turns but supercoils too. One might ask, therefore, how the removal of helical turns and supercoils are connected. The linking number, the sum of the helical turns and the (negative) supercoils in a closed DNA domain, must be constant. Therefore, if the number of the helical turns decreases, the number of supercoils (in absolute number) must decrease too to keep the

linking number constant (Marincs and White 1996). This means that in the *noc* operon the numbers of negative supercoils are higher in the repressed than in the activated state, i.e. during activation the *noc* promoter region becomes less supercoiled or more relaxed in other words. Thus, removal of helical turns and supercoils occur simultaneously during activation of *nocP*.

The mechanistic model described for the regulation of gene expression in the *noc* operon of pTiT37 (Figure 15 and Marincs and White 1996), is in good agreement with those theoretical conceptions that modulation of DNA supercoiling over a distance and B- to Z-DNA transitions might be able to turn genes on and off (Sinden 1994b).

It is worth to note that, according to my knowledge, no similar complex regulation was described in the literature to date. However, components similar to of those of the *noc* regulatory mechanism can be found or were modelled in other prokaryotic systems. For example, in the mercury resistance (*mer*) operon a putative Z-DNA forming sequence overlaps with an operator for the MerR regulatory protein and this operator is very similar to the *noc* operator (Ansari et al. 1992). Some years ago it was also demonstrated in a bacterial plasmid model system that sequence-specific DNA-binding proteins affect transcription-coupled supercoiling in the presence of their operator (Leng and McMacken 2002).

3.8. The biological sense of the described regulatory model

One might ask, that why a simple-looking biochemical mechanism, like nopaline uptake, is regulated in such a relatively complicated manner as I have described above.

I have to acknowledge that although my results were quite unique at the time when they were published, there are more and more knowledge about the involvement of the different components of the above described mechanism in regulation of gene expression, and I see these new results as evidences for the validity of my concept.

For example, a vast number of results have been published about the involvement of DNA superhelicity and Z-DNA formation in regulation of gene expression, since the regulatory model for the *noc* operon was published, and below I highlight some of these results.

Now we know that quite a large number of bacterial genes are regulated by the level of DNA superhelicity and that particular promoters can be turned on and off at a certain level of superhelical density (Lim et al. 2003; Peter et al. 2004; Travers and Muskhelishvili 2005).

Sequences in the upstream region of promoters in eukaryotes, are in Z-DNA form when the promoter is transcribed; however it is not know yet that the Z-DNA formation is the prerequisite or consequence of transcription (Rich and Zhang 2003).

It was also demonstrated that transcriptional repression and activation are effected by superhelicity induced B- to Z-DNA transition in the upstream region of the regulated promoter (Sheridan et al. 2001). Moreover, transcriptional coupling was demonstrated in divergently transcribed promoters where negative DNA supercoiling is dependent on transcription-generated superhelicity and is proportional to promoter strength and transcript length (Opel and Hatfield 2001).

Both Z-DNA binding and sequence-specific DNA-binding proteins were demonstrated to be involved in supercoiling-mediated gene expression (Leng and McMacken 2002; Oh et al. 2002).

Divergently transcribed genes, such as the *nocR-nocP* arrangement, are very common in bacteria. About 40% of all transcription units are transcribed from divergent promoters and a large proportion of such operons contain a regulatory gene (Opel et al. 2001). In the *noc* region, one gene, *nocR* is encoding a regulator gene, while the other gene, the regulated *nocP* encodes a putative ABC-transporter protein, which is very likely involved in the uptake of nopaline (Marincs and White, unpublished result).

Based on a number of pieces of experimental evidences, it is clear now that transcription-mediated DNA supercoiling is involved in the regulation of expression of divergently transcribed genes. One enzyme involved in changing and maintaining DNA superhelicity is gyrase whose activity depends on cellular [ATP/ADP] ratio (Opel et al. 2001). Cellular [ATP/ADP] ratio is correlated with growth conditions, and consequently the level of DNA supercoiling also depends on those factors. When cellular [ATP/ADP] ratio is high then negative supercoil level is also high, and when [ATP/ADP] ratio is low then negative supercoil level is also low. For example, in non-growing cells the [ATP/ADP] ratio is low and superhelical density decreases as a consequence (Opel et al. 2001).

How does this apply to the nopaline uptake of *A. tumefaciens*?

The bacterium is living in a relatively nutrient-free environment, where the cellular level of the [ATP/ADP] ratio and consequently the cellular superhelical level should be low, and it was described above that at decreased local superhelicity the *nocP* promoter which controls genes involved in nopaline uptake is expressed stronger than the basal level.

In vitro, I have found that in a minimal medium lacking nopaline and any other carbon source and in which bacterial ATP level decreased, expression of the *nocP* promoter was about

2.5-fold higher than the basal level. Under similar conditions in nature, when nopaline is absent and cells are starving (i.e. the cellular [ATP/ADP] ratio is low), the nopaline uptake proteins are expressed at a low but definite level by a mechanism in which superhelicity is reduced and consequently no extra energy is consumed for maintaining local supercoils. This makes biological sense in that way that in the starving bacteria the uptake proteins are available even before nopaline is actually synthesised in the incited tumour. The presence of the uptake proteins ensures that when nopaline becomes available due to its synthesis in the tumour, bacteria are ready to take it up immediately and use it as carbon, nitrogen and energy sources.

4. Conclusions

In this thesis and the supplemental publications, the structural and functional characterisation of the nopaline catabolism operon of the *A. tumefaciens* Ti plasmid pTiT37 and a possible mechanism for regulation of gene expression in that operon are described.

The results are summarised below:

- The predicted regulatory gene, *nocR*, and part of the divergently transcribed *nocP* gene of pTiT37 were sequenced.
- An ORF, a promoter and a SD sequence for *nocR* were identified and it was demonstrated that the deduced protein sequence of that ORF is homologous to members of the LysR family of prokaryotic activator proteins, and has a helix-turn-helix DNA-binding motif near its N-terminal. The NocR protein was expressed in *E. coli* and it was demonstrated that the molecular mass and the first ten amino acids of the expressed protein are identical to of those of the deduced protein sequence.
- It was demonstrated that the NocR protein has distinct regulatory functions *in vivo*. It is a repressor for its own synthesis and a dual function repressor/activator for the expression of the *nocP* gene.
- The regulated *nocP* promoter was located by functional and sequence analysis. Its structural characters make it probable that the *nocP* promoter is supercoiling responsive.
- The binding site of NocR was identified in the *nocR-nocP* intervening region. NocR binds to its own promoter under both repressive and inductive conditions. Nopaline, the inducer of *nocP* expression has no effect on binding, but induces conformational changes in the NocR-DNA complex.
- In the region protected by NocR, three overlapping sequences with putative regulatory functions were identified. These are: a $CATGN_4CATG$ tandem repeat, an 18 bp alternating purine-pyrimidine putative Z-DNA forming sequence and a putative gyrase recognition site.
- The $CATGN_4CATG$ sequence was proven to serve as an operator for NocR both *in vitro* and in vivo. It was established that this is the only operator in the *nocR-nocP* promoter region, occupying a remote position relative to the regulated *nocP* promoter.
- The *noc* operator was demonstrated to modulate DNA supercoiling in vivo.
- It was demonstrated that in addition to the NocR protein, the *noc* operator and divergent transcription also play an independent role in the expression of the *nocP* gene in vivo.

- Based on the experimental evidences it was suggested that local DNA supercoiling might have a role in the NocR-mediated expression of the *nocP* gene. It was also suggested that the *noc* operator kept in or converts into Z-DNA by the NocR protein might be the driving force behind activation of the *nocP* promoter.
- A mechanistic model describing gene regulation in the *noc* operon of pTiT37 was presented. According to this model, the *nocP* gene is repressed in the absence of nopaline because its sub-optimally spaced -10 and -35 hexamers are kept in that position by a certain level of local supercoiling maintained by transcription of the divergent *nocR* gene, binding of NocR to the *noc* operator and possibly by a gyrase. In the presence of nopaline, a structural change occurs in the NocR protein-*noc* operator complex, which can be either a B- to Z-DNA transition or an extension of a Z-DNA stretch This structural change results in removals of both helical turns and local supercoils in the *noc* domain bringing the hexamers of the *nocP* promoter into the optimal spatial conformation relative to each other, thus activating expression of *nocP*.
- This model, based on experimental findings in a natural system supports theoretical conceptions suggesting that regulation of gene expression is possible through modulating the structure of DNA over a distance by a regulatory protein bound to a single operator and that B- to Z-DNA transitions might turn genes on and turn off.

5. Literature

The thesis is based on the following publication:

Marincs, F. and White, D.W.R. (1993) Nopaline causes a conformational change in the NocR regulatory protein-*nocR* promoter complex of *Agrobacterium tumefaciens* Ti plasmid pTiT37. *Mol. Gen. Genet.* **24:** 65-72.

Marincs, F. and White, D.W.R. (1994) The NocR repressor-activator protein regulates expression of the *nocB* and *nocR* genes of *Agrobacterium tumefaciens*. *Mol. Gen. Genet.* **244:** 367-73.

Marincs, F. and White, D.W.R. (1995) Divergent transcription and a remote operator play a role in control of expression of a nopaline catabolism promoter in *Agrobacterium tumefaciens*. *J. Biol. Chem.* **270:** 12339-12342.

Marincs, F. and White, D.W.R. (1996) Regulation of gene expression at a distance: the hypothetical role of regulatory protein-mediated topological changes of DNA. *FEBS Lett.* **382:** 1-5.

Cited literature:

Adhya, S. and Garges, S. (1990) Positive control. *J. Biol. Chem.* **265:** 10797-10800.

Ansari, A.Z., Chael, M.L. and O'Halloran, T.V. (1992) Allosteric underwinding of DNA is a critical step in positive control of transcription by Hg-MerR. *Nature* **355:** 87-89.

Babu, M.M. and Teichmann, S:A: (2003) Evolution of transcription factors and the gene regulatory network in *Escherichia coli*. *Nucleic Acids Res.* **31:** 1234-1244.

Balke, V.L. and Gralla, J.D. (1987) Changes in the linking number of supercoiled DNA accompany growth transitions in *Escherichia coli*. *J. Bacteriol.* **169:** 4499-4506.

Beauchamp, C.J., Chilton, W.S., Dion, P. and Antoun, H. (1990) Fungal catabolism of crown gall opines. *Appl. Environ. Microbiol.* **56:** 150-155.

Beck, C.F. and Warren, R.A.J. (1988) Divergent promoters, a common from of gene organisation. *Microbiol. Rev.* **52:** 318-326.

Bergeron, J., MacLeod, R.A. and Dion, P. (1990) Specificity of octopine uptake by *Rhizobium* and *Pseudomonas* strains. *Appl. Environ. Microbiol.* **56:** 1453-1458.

Bomhoff, G., Klapwijk, P.M., Kester, H.C.M., Schilperoort, R.A., Hernalsteens, J.P. and Schell, J. (1976) Octopine and nopaline synthesis and breakdown genetically controlled by a plasmid of *Agrobacterium tumefaciens*. *Mol. Gen. Genet.* **145:** 177-181.

Brencic, A., Angert, E.R. and Winans, S.C. (2005) Unwounded plants elicit *Agrobacterium vir* gene induction and T-DNA transfer: transformed plant cells produce opines yet are tumour free. *Mol. Microbiol.* **57:** 1522-1531.

Brencic, A. and Winans, S.C. (2005) Detection of and response to signals involved in host-microbe interactions by plant-associated bacteria. *Microbiol. Mol. Biol. Rev.* **69:** 155-194.

Brennan, R.G. and Matthews, B.W. (1989) The helix-turn-helix DNA binding motif. *J. Biol. Chem.* **264:** 1903-1906.

Brill, S.J. and Sternglanz, R. (1988) Transcription-dependent DNA supercoiling in yeast DNA topoisomerase mutants. *Cell* **54:** 403-411.

Champoux, J.J. (2001) DNA topoisomerases: structure, function, and mechanism. *Annu. Rev. Biochem.* **70:** 369-413.

Christman, M.F., Storz, G. and Ames, B.N. (1989) OxyR, a positive regulator of hydrogen peroxide-inducible genes in Escherichia coli and Salmonella typhimurium, is homologous to a family of bacterial regulatory proteins. *Proc. Natl. Acad. Sci. USA* **86:** 3484-3488.

Collado-Vides, J., Magasnik, B. and Gralla, J.D. (1991) Control site location and transcriptional regulation in *Escherichia coli*. *Microbiol. Rev.* **55:** 371-394.

Condee, C.W. and Summers, A.O. (1992) A *mer-lux* transcriptional fusion for real-time examination of in vivo gene expression kinetics and promoters response to altered superhelicity. *J. Bacteriol.* **174:** 8094-8101.

Depicker, A., Stachel, S., Dhaese, P., Zambryski, P. and Goodman. H.M. (1982) Nopaline synthase: transcript mapping and DNA sequence. *J. Mol. Appl. Genet.* **1:** 561-573.

Deng, S., Stein, R.A. and Higgins, N.P. (2005) Organization of supercoil domains and their reorganization by transcription. *Mol. Microbiol.* **57:** 1511-1521.

Dessaux, Y., Petit, A. and Tempe, J. (1993) Chemistry and biochemistry of opines, chemical mediators of parasitism. *Phytochemistry* **34:** 31-38.

Figueroa, N. and Bossi, L. (1988) Transcription induces gyration of the DNA template in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **85:** 9416-9420.

Gelvin, S. B. (2003) Agrobacterium-mediated plant transformation: the biology behind the “gene-jockeying” tool. *Microbiol. Mol. Biol. Rev.* **67:** 16-37.

Goethals, K., Van Montagu, M. and Holsters, M. (1992) Conserved motifs in a divergent *nod* box of *Azorhizobium caulinodans* ORS571 reveal a common structure in promoters regulated by LysR-type proteins. *Proc. Natl. Acad. Sci. USA* **89:** 1646-1650.

Goodrich, J.A. and McClure, W.R. (1991) Competing promoters in prokaryotic transcription. *Trends Biochem. Sci.* **16:** 394-397.

Gralla, J.D. (1991) Transcription control – lessons from an *E. coli* promoter data base. *Cell* **66:** 415-418.

Greer, L.F. and Szalay, A.A. (2002) Imaging of light emission from the expression of luciferases in living cells and organisms: a review. *Luminescence* **17:** 43-74.

Guyon, P., Petit, A., Tempé, J. and Dessaux, Y. (1993) Transformed plants producing opines specifically promote growth of opine-degrading agrobacteria. *Mol. Plant-Microbe Interact.* **6:** 92-98.

Ha, S.C., Lowenhaupt, K., Rich, A., Kim, Y.G. and Kim, K.K. (2005) Crystal structure of a junction between B-DNA and Z-DNA reveals two extruded bases. *Nature* **437:** 1183-1186.

Habeeb, L.F., Wang, L. and Winans, S.C. (1991) Transcription of the octopine catabolism operon of the *Agrobacterium* tumor-inducing plasmid pTiA6 is activated by a LysR-type regulatory protein. *Mol. Plant-Microbe Interact.* **4:** 379-385.

Harrison, S.C. (1991) A structural taxonomy of DNA-binding domains. *Nature* **353:** 715-719.

Henikoff, S., Haughn, G.W., Calvo, J.M. and Wallace, J.C. (1988) A large family of bacterial activator proteins. *Proc. Natl. Acad. Sci. USA* **85:** 6602-6606.

Herbert, A. and Rich, A. (1999) Left-handed Z-DNA: structure and function. *Genetica* **106:** 37-47.

Holsters, M., Silva, B., Van Vliet, F., Genetello, C., De Block, M., Dhaese, P., Depicker, A., Inzé, D., Engler, G., Villaroel, R., Van Montagu, M. and Schell, J. (1980) The functional organisation of the nopaline *A. tumefaciens* plasmid pTiC58. *Plasmid* **3:** 212-230.

Holtham, C.A.M., Jumel, K., Miller, C.M., Harding, S.E., Baumberg, S. and Stockley, P.G. (1999) Probing activation of the prokaryotic arginine transcriptional regulator using chimeric proteins. *J. Mol. Biol.* **289:** 707-727.

Hsieh, L.S., Rouviere-Yaniv, J. and Drlica, K. (1991) Bacterial DNA supercoiling and [ATP]/[ADP] ratio: changes associated with salt shock. *J. Bacteriol.* **173:** 3914-3917.

Ishihama, A. (1993) Protein-protein communication within the transcription apparatus. *J. Bacteriol.* **175:** 2483-2489.

Jefferson, R.A., Kavanagh, T.A. and Bevan, M.W. (1987) GUS fusions, β-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J.* **6:** 3901-3907.

Kaphammer, B. and Olsen, R.H. (1990) Cloning and characterisation of TfdS, the repressor-activator of *tfdB*, from the 2,4-dichlorophenoxyacetic acid catabolic plasmid pJP4. *J. Bacteriol.* **172:** 5856-5862.

Kim, K-S. and Farrand, S.K. (1996) Ti plasmid-encoded genes responsible for catabolism of the crown gall opine mannopine by *Agrobacterium tumefaciens* are homologs of the T-region genes responsible for synthesis of this opine by the plant tumor. *J. Bacteriol*. **178:** 3275-3284.

Kim, H. and Farrand, S.K. (1998) Opine catabolic loci from Agrobacterium plasmids confer chemotaxis to their cognate substrates. *Mol. Plant-Microbe Interact.* **11:** 131-143.

Kim, K-S., Baek, C-H., Lee, J.K., Yang, J.M. and Farrand, S.K. (2001) Intracellular accumulation of mannopine an opine produced by crown gall tumors, transiently inhibits growth of *Agrobacterium tumefaciens*. *Mol. Plant-Microbe Interact.* **14:** 793-803.

Klapwijk, P.M., Oudshoorn, M. and Schilperoort, R.A. (1977) Inducible permease involved in the uptake of octopine, lysopine and octopinic acid by *Agrobacterium tumefaciens* strains carrying virulence-associated plasmids. *J. Gen. Microbiol.* **102:** 1-11.

Kölling, R. and Lother, H. (1985) AsnC: an autogenously regulated activator of asparagine synthetase A transcription in *Escherichia coli*. *J. Bacteriol.* **164:** 310-315.

Kredich, N.M. (1992) The molecular basis for the positive regulation of *cys* promoters in *Salmonella typhimurium* and *Escherichia coli*. *Mol. Microbiol.* **6:** 2747-2753.

Krishnan, M., Burgner, J.W., Chilton, W.S. and Gelvin, S.B. (1991) Transport of nonmetabolizable opines by *Agrobacterium tumefaciens*. *J. Bacteriol.* **173:** 903-905.

Ladant, D. and Karimova, G. (2000) Genetic systems for analyzing protein-protein interactions in bacteria. *Res. Microbiol.* **151:** 711-720.

Leng, F. and McMacken, R. (2002) Potent stimulation of transcription-coupled DNA supercoiling by sequence-specific DNA-binding proteins. *Proc. Natl. Acad. Sci. USA* **99:** 9139-9144.

Lewin, B. (1990) Genes IV. Oxford University Press, New York and Cell Press, Cambridge, Mass., pp 257-259.

Lim, H.M., Lewis, D.E.A., Lee, H.J., Liu, M. and Adhya, S. (2003) Effect of varying the supercoiling of DNA on transcription and its regulation.

Lisser, S. and Margalit, H. (1993) Compilation of *E. coli* mRNA promoter sequences. *Nucleic Acids Res.* **21:** 1507-1516.

Liu, P. and Nester, E.W. (2006) Indoleacetic acid, a product of transferred DNA, inhibits *vir* gene expression and growth of *Agrobacterium tumefaciens* C58. *Proc. Natl. Acad. Sci. USA* **103:** 4658-4662.

Liu, L.F. and Wang, J.C. (1987) Supercoiling of the DNA template during transcription. *Proc. Natl. Acad. Sci. USA* **84:** 7024-7027.

Lockshon, D. and Morris, D.R. (1985) Sites of reaction of *Escherichia coli* gyrase on pBR322 in vivo as revealed by oxolinic acid-induced plasmid linearization. *J. Mol. Biol.* **181:** 63-74.

Marincs, F. and White, D.W.R. (1993) Nopaline causes a conformational change in the NocR regulatory protein-*nocR* promoter complex of *Agrobacterium tumefaciens* Ti plasmid pTiT37. *Mol. Gen. Genet.* **24:** 65-72.

Marincs, F. and White, D.W.R. (1994) The NocR repressor-activator protein regulates expression of the *nocB* and *nocR* genes of *Agrobacterium tumefaciens*. *Mol. Gen. Genet.* **244:** 367-73.

Marincs, F. and White, D.W.R. (1995) Divergent transcription and a remote operator play a role in control of expression of a nopaline catabolism promoter in *Agrobacterium tumefaciens*. *J. Biol. Chem.* **270:** 12339-12342.

Marincs, F. and White, D.W.R. (1996) Regulation of gene expression at a distance: the hypothetical role of regulatory protein-mediated topological changes of DNA. *FEBS Lett.* **382:** 1-5.

Matthews, K.S. (1992) DNA looping. *Microbiol. Rev.* **56:** 123-136.

Moore, L.W., Chilton, W.S. and Canfield, M.L. (1997) Diversity of opines and opine-catabolizing bacteria isolated from naturally occurring crown gall tumors. *Appl. Environ. Microbiol.* **63:** 201-207.

Nautiyal, C.S. and Dion, P. (1990) Characterization of the opine-utilizing microflora associated with samples of soils and plants. *Appl. Environ. Microbiol.* **56:** 2576-2579.

Oger., P. and Farrand, S. K. (2002) Two opines control conjugal transfer of an *Agrobacterium* plasmid by regulating expression of separate copies of the quorum-sensing activator gene *traR*. *J. Bacteriol*. **184:** 1121-1131.

Oger, P., Petit, A. and Dessaux, Y. (1997) Genetically engineered plants producing opines alter their biological environment. *Nat. Biotechnol.* **15:** 369-372.

Oh, D.B., Kim, Y.G. and Rich, A. (2002) Z-DNA-binding proteins can act as potent effectors of gene expression *in vivo*. *Proc. Natl. Acad. Sci. USA* **99:** 16666-16671.

Opel, M.L., Arfin, S.M. and Hatfield, G.W. (2001) The effects of DNA supercoiling on the expression of operons of the *ilv* regulon of *Escherichia coli* suggest a physiological rationale for divergently transcribed operons. *Mol. Microbiol.* **39:** 1109-1115.

Opel, M.L. and Hatfield, G.W. (2001) DNA supercoiling-dependent transcriptional coupling between the divergently transcribed promoters of the *ilvYC* operon of *Escherichia coli* is proportional to promoter strengths and transcript lengths. *Mol. Microbiol.* **39:** 191-198.

Pappas, K.M. and Winans, S.C. (2003) A LuxR-type regulator from *Agrobacterium tumefaciens* elevates Ti plasmid copy number by activating transcription of plasmid replication genes. *Mol. Microbiol.* **48:** 1059-1073.

Peter, B.J., Arsuaga, J., Breier, A.M., Khodursky, A.B., Brown, P.O. and Cozzarelli, N.R. (2004) Genomic transcriptional response to loss of chromosomal supercoiling in *Escherichia coli*. *Genome Biol.* **5:** R87.

Pérez-Martin, J. and de Lorenzo, V. (1995) The sigma 54-dependent promoter Ps of the TOL plasmid of *Pseudomonas putida* requires HU for transcriptional activation in vivo by XylR. *J. Bacteriol.* **177:** 3758-3763.

Pérez-Rueda, E. and Collado-Vides, J. (2000) The repertoire of DNA-binding transcriptional regulators in *Escherichia coli* K-12. *Nucleic Acids Res.* **28:** 1838-1847.

Pérez-Rueda, E. and Collado-Vides, J. (2001) Common history at the origin of the position-function correlation in transcriptional regulators in archaea and bacteria. *J. Mol. Evol.* **53:** 172-179.

Pruss, G.J. and Drlica, K. (1989) DNA supercoiling and prokaryotic transcription. *Cell* **56:** 521-523.

Ptashne, M. (2005) Regulation of transcription: from lambda to eukaryotes. *TRENDS Biochem. Sci.* **30:** 275-279.

Rich, A. and Zhang, S. (2003) Z-DNA: the long road to biological function. *Nature Rev. Genet.* **4:** 566-572.

Rhodius, V.A. and Busby, S.J.W. (1998) Positive activation of gene expression. *Curr. Opinion Microbiol.* **1:**152-159.

Rojo, F. (1999) Repression of transcription initiation in Bacteria. *J. Bacteriol.* **181:** 2987-2991.

Rosenberg, M. and Court, D. (1979) Regulatory sequences involved in the promotion and termination of RNA transcription. *Annu. Rev. Genet.* **13:** 319-353.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular cloning: A laboratory manual. 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA.

Savka, M.A. and Farrand, S.K. (1997) Modification of rhizobacterial populations by engineering bacterial utilization of a novel plant-produced resource. *Nat. Biotechnol.* **15:** 363-368.

Schardl, C.L. and Kado, C.J. (1983) Ti plasmid and chromosomal ornithine catabolism genes of *Agrobacterium tumefaciens* C58. *J. Bacteriol.* **155:** 196-202.

Schell, M.A. (1993) Molecular biology of the LysR family of transcriptional regulators. *Annu. Rev. Microbiol.* **47:** 597-626.

Schell , M.A. and Faris, E. (1987) Transcriptional regulation of the *nah* and *sal* naphthalane degradation operons of plasmid NAH7 of *P. putida*. In: Reznikoff, W.S., Burgess, R., Dahlberg, J., Gross, C. and Record, M. (eds) RNA polymerase and the regulation of transcription. Elsevier, New York, pp 455-458.

Schleif, R. (1992) DNA looping. *Annu. Rev. Biochem.* **61:** 199-223.

Semsey, S., Virnik, K. and Adhya, S. (2005) A gamut of loops: meandering DNA. *TRENDS Biochem. Sci.* **30:** 334-341.

Sheridan, S.D., Opel, M.L. and Hatfield, G.W. (2001) Activation and repression of transcription initiation by a distant structural transition. *Mol. Microbiol.* **40:** 684-690.

Sinden, R.R. (1994a) DNA structure and function. Academic Press, San Diego, CA, USA. p.190.

Sinden, R.R. (1994b) DNA structure and function. Academic Press, San Diego, CA, USA. p.207-213.

Smith, G.R. (1981) DNA supercoiling: another level for regulating gene expression. *Cell* **24:** 599-600.

Stachel, S.E., An, G., Flores, C. and Nester, E.W. (1985) A Tn3 *lac* transposon for the random genertaion of β-galactosidase gene fusions: application to the analysis of gene expression in *Agrobacterium*. *EMBO J.* **4:** 891-898.

Su, W., Porter, S., Kustu, S. and Echols, H. (1991) DNA-looping and enhancer activity: Association between DNA-bound NtrC activator and RNA polymerase at the bacterial *glnA* promoter *Proc. Natl. Acad. Sci. USA* **87:** 5504-5508.

Tempé, J. and Petit, A (1982) Opine utilisation by *Agrobacterium*. In: Kahl, G. and Shell, J.S. (eds) Molecular biology of plant tumors. Academic Press, Orlando, USA, pp 451-459.

Travers A.A. (1989) DNA conformation and protein binding. *Annu. Rev. Biochem.* **58:** 427-452.

Travers, A. and Muskhelishvili, G. (2005) DNA supercoiling – A global transcriptional regulator for enterobacterial growth? *Nature Rev. Microbiol.* **3:** 157-169.

Tremblay, G., Gagliardo, R., Chilton, W.S. and Dion, P. (1987) Diversity among opine-utilizing bacteria: identification of coryneform isolates. *Appl. Environ. Microbiol.* **53:** 1519-1524.

Tsao, Y.P., Wu, H.Y. and Liu, L.F. (1989) Transcription-driven supercoiling of DNA: direct biochemical evidence from in vitro studies. *Cell* **56:** 111-118.

Tzfira, T., Li, J., Lacroix, B. and Citovsky, V. (2004) *Agrobacterium* T-DNA integration: molecules and models. *TRENDS Genet.* **20:** 375-383.

Valdivia, R.H., Wang, L. and Winans, S.C. (1991) Characterization of a putative periplasmic transport system for octopine accumulation encoded by *Agrobacterium tumefaciens* Ti plasmid pTiA6. *J. Bacteriol.* **173:** 6398-6405.

Valentine, L. (2003) *Agrobacterium tumefaciens* and the plant: The David and Goliath of modern genetics. *Plant Physiol.* **133:** 948-955.

Van Larebeke, N., Engler, G., Holsters, M., Van den Elsacker, S., Iaenen, I., Schilperoort, R.A. and Schell, J. (1974) Large plasmid in *Agrobacterium tumefaciens* essential for crown gall inducing ability. *Nature* **252:** 169-170.

Veluthambi, K., Krishnan, M., Gould, J.H., Smith, R.H. and Gelvin, S.B. (1989) Opines stimulate induction of the *vir* genes of the *Agrobacterium tumefaciens* Ti plasmid. *J. Bacteriol.* **171:** 3696-3703.

Vogel, C., Bashton. M., Kerrison, N.C., Chothia, C. and Teichmann, S.A. (2004) Structure, function and evolution of multidomain proteins. *Curr. Opinion Struct. Biol.* **14:** 208-216.

Von Lintig, J., Zanker, H. and Schröder, J. (1991) Positive regulators of opine-inducible promoters in the nopaline and octopine catabolism regions of Ti plasmids. *Mol. Plant-Microbe Interact.* **4:** 370-378.

Von Lintig, J., Zanker, H. and Schröder, J. (1994) Opine-regulated promoters and LsyR-type regulators in the nopaline (*noc*) and Octopine (*occ*) regions of Ti plasmids of *Agrobacterium tumefaciens*. *J. Bacteriol.* **176:** 495-503.

Zanker, H., Lurz, G., Langridge, U., Langridge, P., Kreusch, D. and Schröder, J. (1994) Octopine and nopaline oxidases from Ti plasmids of *Agrobacterium tumefaciens*: Molecular analysis, relationship, and functional characterization. *J. Bacteriol.* **176:** 4511-4517.

Zaim, J. and Kierzek, A.M. (2003) The structure of full-length LysR-type transcriptional regultors. Modelling of the full-length OxyR transcription factor dimer. *Nucleic Acids Res.* **31:** 1444-1454.

Zupan J., Muth, T.R., Draper, O. and Zambryski P. (2000) The transfer of DNA from *Agrobacterium tumefaciens* into plants: a feast of fundamental insights. *The Plant J.* **23:** 11-28.

Zhu, J., Oger, P.M., Schrammeijer, B., Hooykaas, P.J.J., Farrand, S.K. and Winans, S.C. (2000) The bases of crown gall tumorigenesis. *J. Bact.* **182:** 3885-3895.

Wang, J.C. and Giaever, G.N. (1988) Action at a distance along a DNA. *Science* **240:** 300-304.

Wang, J.Y. and Syvanen, M. (1992) DNA twist as a transcriptional sensor for environmental changes. *Mol. Microbiol.* **6:** 1861-1866.

Wang, L., Helmann, J.D. and Winans, S.C. (1992) The *A. tumefaciens* transcriptional activator OccR causes a bend at a target promoter, which is partially relaxed by a plant tumor metabolite. *Cell* **69:** 659-667.

Weickert, M.J. and Adhya, S. (1992) A family of bacterial regulators are homologous to Gal and Lac repressors. *J. Biol. Chem.* **267:** 15869-15874.

Xu, H. and Hoover, T.R. (2001) Transcriptional regulation at a distance in bacteria. *Curr. Opinion Microbiol.* **4:** 138-144.

Other publications from the topic of the thesis:

White, D.W.R., Pritchard, M. and Marincs, F. (1990): Genetic and sequence analysis of *Agrobacterium* opine catabolism genes. Molecular Genetics Symposia, Palmerston North, New Zealand.

Marincs, F. and White, D.W.R. (1991): Regulation of *Agrobacterium* opine catabolism gene expression. 1st Annual Queenstown Molecular Biology Meeting, Queenstown, New Zealand.

White, D.W.R. and Marincs, F. (1991): Molecular interactions in the activation of *Agrobacterium* opine catabolism gene expression. Australasian Gene Mapping and Molecular Genetics Symposium, Dunedin, New Zealand.

Marincs, F. and White, D.W.R. (1992): Gene manipulation to develop microbes as biosensors. 1st New Zealand Biosensor Symposium, Palmerston North, New Zealand.

Marincs, F. and White, D.W.R. (1992): DNA topology controls expression of the *nocB* gene in *Agrobacterium tumefaciens*. Meeting on Molecular Genetics of Bacteria and Phages, Cold Spring Harbor, NY, USA.

Marincs, F. and White, D.W.R. (1992): DNA topology controls expression of the *nocB* gene in *Agrobacterium tumefaciens*. 2nd Annual Queenstown Molecular Biology Meeting, Queenstown, New Zealand.

Marincs, F. and White, D.W.R. (1993): Regulation of nopaline uptake in *Agrobacterium tumefaciens*: in vivo and in vitro studies. Joint Annual Conference of the New Zealand Microbiological Society and the New Zealand Society for Biochemistry and Molecular Biology. Palmerston North, New Zealand.

Marincs, F. and White, D.W.R. (1994): Effect of upstream sequences on expression of a bacterial promoter. 4th Queenstown Molecular Biology Meeting, Queenstown, New Zealand.

Marincs, F. and White, D.W.R. (1994): Regulation of the nopaline catabolism genes of pTiT37 in *Agrobacterium tumefaciens*. 7th International Symposia on Molecular Plant-Microbe Interactions, Edinburgh, Scotland.

Marincs, F., Dudas, B. and White, D.W.R. (1994): Activation of an *Agrobacterium* nopaline catabolism promoter in planta. 4th Queenstown Molecular Biology Meeting, Queenstown, New Zealand.

Other publications of the author:

Dallmann, G., Marincs, F., Papp, P., Gaszner, M. and Orosz L. (1991) The isolated N-terminal DNA binding domain of the c repressor of bacteriophage 16-3 is functional in DNA binding in vivo and in vitro. *Mol. Gen. Genet.* **227:** 106-112.

Marincs, F. and White, D.W.R. (1994) Immobilization of *Escherichia coli* expressing the *lux* genes of *Xenorhabdus luminescens*. *Appl. Environ. Microbiol.* **60:** 3862-3863.

Voisey, C.R. and Marincs, F. (1998) Elimination of internal restriction enzyme sites from a bacterial luminescence (*luxCDABE*) operon. *Biotechniques* **24:** 56-58.

Marincs F. (2000) On-line monitoring of growth of *Escherichia coli* in batch cultures by bioluminescence. *Appl. Microbiol. Biotechnol.* **53:** 536-541.

Marincs, F., McDowall, K.J. and Stockley, P.G. (2004): A combined *in vitro* transposon-*in vivo* recombination mutagenesis method to knock out genes in *Escherichia coli*. *Amer. Biotechnol. Lab.* **22:** 8-10.

Marincs, F., Manfield, I.W., Stead, J.A., McDowall, K.J. and Stockley, P.G. (2006) Transcript analysis reveals an extended regulon and the importance of protein-protein co-operativity for the *Escherichia coli* methionine repressor. *Biochem. J.* **396:** 227-234.

6. Összefoglaló

a

„Regulation of nopaline catabolism in *Agrobacterium tumefaciens*”

című PhD disszertációhoz

Szerző: Dr Marincs Ferenc

Baktériumokban a gének megnyilvánulása különböző szinteken szabályozódik, amelyek közül a legelső a DNS kötő fehérjék által szabályozott átírás vagy más néven transzkripció. Ezek a fehérjék a szabályozott promoterek környezetében lévő operátor szekvenciákhoz kötődve, kis molekulák hatására vagy megakadályozzák (represszálják) vagy elősegítik (aktiválják) az adott gének megnyilvánulását és ezáltal biztosítják a géntermékek jelenlétét a sejtek biokémiai folyamataihoz.

A génmegnyilvánulás szabályozásának különböző mechanizmusai vannak, amelyek során a szabályozott promoter régiókban DNS-fehérje és fehérje-fehérje kölcsönhatások és bizonyos DNS valamint fehérjeszintű szerkezeti változások egyaránt részt vesznek.

Az *Agrobacterium tumefaciens* egy növénypatogén baktérium, amely tumort hoz létre magasabb rendű növényeken azáltal, hogy nagyméretű, úgy nevezett tumor-indukáló (Ti) plazmidjának egy bizonyos szegmentjét (T-DNS) átviszi a növényi sejtekbe, ahol az beépül a genomba. A T-DNS integrációját követően, a beépült gének megnyilvánulása növényi hormonok túltermelődését eredményezi, ami a bakteriális génekkel transzformált növényi sejtek kontrolálhatatlan osztódásához, és így a tumor kialakulásához vezet. Ezen túlmenően, a növényi genomba integrálódott bakteriális gének speciális vegyületek, un. opinok termelését is irányítják a tumor sejtjeiben. Az opinokat a tumort indukáló baktérium képes lebontani és kizárólagos szén-, nitrogén- és energia-forrásként hasznosítani. Az opinok lebontását olyan a

Ti plazmidon található gének vezérlik, amelyek nem részei a növényi genomba integrálódó T-DNS régiónak. Ezzel a tumor indukáló mechanizmussal a baktérium egy olyan különleges környezetet létesít ahol a faj túlélése és szaporodása biztosított.

Jelen disszertációban egy, az *Agrobacterium tumefaciens* pTiT37-es Ti plazmidjára jellemző opin, a nopaline lebontó operonjának molekuláris vizsgálatát és az abban megfigyelt új típusú szabályozó mechanizmust írom le.

Munkámat megelőzően, laboratóriumunkban a nopaline lebontásban szerepet játszó géneket transzpozon mutagenezissel azonosították. Ezek közül kettő, a *nocP* és *nocR* gének, egymással ellentétes irányban íródnak át. A *nocP* transzkripciója nopalinnal indukálható, míg a *nocR* transzkripciója konstitutív. Meghatároztam a *nocP-nocR* régió DNS szekvenciáját, amelynek analízisével arra a következtetésre jutottam, hogy a *nocR* gén feltehetően egy DNS-kötő, a bakteriális LysR típusú aktivátor fehérjék családjába tartozó fehérjét, a *nocP* pedig egy valószínűleg ABC-tipusú, a nopaline felvételében nagy valószínűséggel szerepet játszó transzport fehérjét kódol.

Jelző gének *in vivo* használatával kimutattam, hogy a NocR fehérje egy kettős funkciójú szabályozó fehérje, amely represszálja a saját fehérjét kódoló *nocR* gén átíródását és aktiválja az ellentétes irányban átíródó *nocP* gén megnyilvánulását. A *nocP* és *nocR* gének promotereit szekvencia és funkcionális analízissel határoztam meg, amiből kiderült, hogy a két promoter nem átfedő és egy 131 bp hosszú, ismeretlen funkciójú rész választja el a két promoter -35-ös elemeit. Mindkét promoter nagymértékben homológ az Escherichia coli σ^{70} promoterek konszenzus szekvenciájával. Ezen túlmenően a *nocP* promoter olyan sajátosságokat is mutat, amelyek a DNS szuperspirál szintre érzékeny promoterekre jellemzőek. A NocR fehérjét Escherichia coliban termeltettem és kimutattam, hogy a fehérje alegység molekulatömege

31.5 kDa. A termelt és a nukleinsav szekvencia alapján feltételezett fehérjék N-terminális szekvenciája azonos volt. Gél retardációs és DNaseI footprinting kísérletekkel kimutattam, hogy a NocR fehérje a *nocP* és *nocR* gének kódoló szekvenciája közötti régióhoz kötődik mind nopaline jelenlétében mind annak hiányában. Habár a nopaline nem befolyásolta a NocR fehérje kötődésének erősségét, hatására megváltozott a NocR fehérje-*noc* DNS komplex konformációja. A NocR fehérje kötőhelye egy 72 bp hosszúságú, a *nocR* promotert tartalmazó szekvenciában volt lokalizálható, amely szekvencia a *nocP* promotertől 5' irányban található. Ebben a kötőhelyben egy $CATGN_4CATG$ tandem palindrom szekvenciát azonosítottam és bizonyítottam, hogy ez a szekvencia a NocR fehérje operátoraként működik *in vitro* és *in vivo*.

Az operátor szekvenciával átfedően két érdekes és jellegzetes szekvencia található: egy 18 bp hosszúságú váltakozó purin-pyrimidin szekvencia, amely képes Z-DNS létrehozására és egy lehetséges gyrase (egy olyan enzim, ami a sejtbeli DNS szuperspirál szint beállításában játszik szerepet) felismerő hely. Jelző génekkel történő *in vivo* vizsgálatok igazolták, hogy a *nocR* gén átíródása önmagában is és az operátor szekvenciával kombinálva is, bizonyos fokig megakadályozza a *nocP* gén megnyilvánulását még a NocR fehérje hiányában is. Ez azt mutatja, hogy a két promoter átíródása kapcsolt és a lokális DNS szuperspirál szint szerepet játszik abban a mechanizmusban, amellyel a NocR fehérje szabályozza a *nocP* gén megnyilvánulását. Azt is kimutattam, hogy az operátor szekvencia befolyásolja egy plazmid szuperspirál szintjét *in vitro*, a sejtbeli általános DNS szuperspirál szinttől függően.

Mindezeket az eredményeket összesítve, egy új típusú mechanikus modellt írtam le, amely lehetséges magyarázattal szolgál arra nézve, hogy a NocR fehérje milyen módon szabályozza az Agrobacterium tumefaciens pTiT37-es Ti plazmidjának nopaline lebontó régiójában lévő *nocP* gén megnyilvánulását. A modell szerint, nopaline hiányában a NocR fehérje kötődése a szabályozott *nocP* promoterhez képest távoli pozícióban található *noc* operátorhoz,

kombinálva a *nocR* gén átírása által generált lokális negatív szuperspirálokkal megakadályozza a *nocP* gén megnyilvánulását, mivel ebben a helyzetben a *nocP* promoter elemeinek térbeli viszonya nem megfelelő az átíráshoz. Nopaline jelenlétében, a NocR fehérje-*noc* operator komplex térszerkezete olyan módón változik meg, hogy az operátor vagy átalakul B-formából Z-formába, vagy, ha már eredendően Z-formában volt, akkor a Z-formában lévő rész kiterjed a *nocP* promoter irányába. Ez a térszerkezeti változás a régióban a DNS spirálok és szuperspirálok számának csökkenését eredményezi, ami által a *nocP* promoter elemei az átíráshoz megfelelő térbeli elrendeződésbe kerülnek. Ez a szabályozási mód egy összehangolt, energetikailag előnyös adaptációs és túlélési stratégiát jelent az *Agrobacterium* számára előnytelen életkörülmények között.

Jelen disszertáció négy elsőszerzős, nemzetközi folyóiratokban megjelent közleményen alapszik. Eredményeimet a fenti cikkeken kívül tíz konferencián is bemutattam.

7. Summary

for the PhD thesis

"Regulation of nopaline catabolism in *Agrobacterium tumefaciens*"

Author: Dr Ferenc Marincs

Gene expression in bacteria is regulated at different levels, of which the first one is the regulation of transcription mediated by DNA-binding regulatory proteins. These proteins bind to operator sequences in the vicinity of the regulated promoters, and repress or activate gene expression in response to small effector molecules in order to ensure the correct spatial and temporal presence of gene products for the cellular biochemical processes.

The mechanism of regulation of gene expression by regulatory proteins is implemented by different means in which both DNA-protein and protein-protein interactions in and around the regulated promoter regions, and particular tertiary conformational alterations at the DNA and protein levels may all be involved.

Agrobacterium tumefaciens is a plant pathogenic bacterium which incites tumours on higher plants by transferring a particular portion (the T-DNA) of its large, so-called tumour-inducing (Ti) plasmid into the plant genome. After integration of the T-DNA into the plant chromosomes, expression of genes of the integrated DNA results in unbalanced production of plant hormones, which leads to uncontrolled cell division and consequently to tumour formation. Moreover, the integrated bacterial genes direct the synthesis of specific molecules, collectively called opines, in the tumour cells. Opines are catabolised and utilised as a sole carbon, nitrogen and energy source by the tumour-inducing bacteria. The catabolism of opines is directed by genes located on the Ti plasmids, but these genes are not part of the T-DNA which is transferred into the plant genome.

By this mechanism, the bacteria create a unique niche for themselves, where their survival and multiplication is ensured.

In this dissertation I describe molecular studies performed on the regulation of nopaline catabolism in *A. tumefaciens* and present a novel type of mechanism by which gene expression is regulated in the Ti plasmid pTiT37.

Genes involved in nopaline catabolism had previously been identified in our laboratory by transposon mutagenesis. Two of these genes, *nocR* and *nocP*, were found to be divergently transcribed and expression of *nocP* and *nocR* was nopaline-responsive and nopaline-unresponsive, respectively. The *nocR-nocP* region was sequenced and sequence analysis revealed that the *nocR* gene encodes a putative DNA-binding regulatory protein, which belongs to the family of LysR-type bacterial activator proteins, while the *nocP* gene encodes a putative ABC-type transporter protein which is very likely involved in the uptake of nopaline. In vivo analysis using reporter genes indicated that the NocR protein is a regulatory protein with dual functions, i.e. NocR auto-represses its own synthesis and both represses and activates expression of the divergently transcribed *nocP* gene. The promoters of the *nocR* and *nocP* genes were identified by sequence and functional analyses which revealed that they do not overlap and a 131 bp region of unknown function separates the -35 hexamers of the two promoters. Although both promoters are highly homologous to the consensus *Escherichia coli* σ^{70} promoters, the *nocP* promoter displays features which are characteristic to supercoiling-responsive promoters. The NocR protein was expressed in *E. coli* and it was proven that the molecular mass of its subunit is 31.5 kDa. It was also shown that the N-terminals of the expressed and the predicted proteins are identical. It was shown by gel retardation and DNaseI footprinting assays that the NocR protein binds to the intervening

region between the coding sequences of the *nocR* and *nocP* genes either in the absence or presence of nopaline. Nopaline had no influence on the binding affinity of NocR, but changed the conformation of the NocR-*noc* DNA complex. The binding site of NocR was localised to a 72 bp long DNA sequence in the *nocR* promoter region, upstream of the *nocP* promoter. In the binding site of NocR a $CATGN_4CATG$ tandem palindrome was identified and it was proven both *in vitro* and *in vivo* that this sequence functions as the operator for the NocR protein. Overlapping the operator sequence, two interesting sequence features, an 18 bp alternating purine-pyrimidine sequence able to form Z-DNA, and a putative gyrase (an enzyme involved in adjusting cellular superhelicity) recognition site were found. *In vivo* analyses using reporter genes indicted that transcription of the *nocR* gene on its own and in combination with the operator sequence is able to repress expression of the *nocP* gene to a certain extent even in the absence of the NocR protein. This indicated that the two promoters are transcriptionally coupled and that local superhelicity has a role in the regulatory mechanism by which NocR regulates expression of the *nocP* gene. It was also shown that the operator sequence is able to influence the supercoiling level of a plasmid *in vitro*, depending on the overall level of cellular superhelicity.

Putting these results together, a novel type of mechanistic model was developed to explain the means by which NocR regulates expression of the *nocP* gene in the nopaline catabolism region of the *A. tumefaciens* Ti plasmid pTiT37. According to this model, binding of the NocR regulatory protein to the *noc* operator located in a remote position relative to the regulated *nocP* promoter, in combination with the negative local supercoiling generated by the transcription of the *nocR* gene, represses expression of *nocP*, because in this situation the spatial position of the hexamers of the *nocP* promoter is not favourable for transcription. In the presence of nopaline, the conformation of the NocR protein-*noc* operator complex is altered in such a way that the operator either flips from B- to Z-form or, if it was already in

Z-form in the absence of nopaline, the Z-stretch is expanded towards the *nocP* promoter. This conformational change results in removal of helical turns and local supercoils in the region thus bringing the hexamers of the *nocP* promoter into the correct spatial arrangement for transcription. This means of regulation represents a coordinated, energetically favourable adaptation and survival strategy for *Agrobacterium* living under suboptimal conditions.

This thesis is based on four first-authorship referred publications. The results were also communicated in ten conference presentations.