In search of a biosensor for DNT detection
Studies of inducer response and specificity of DntR

Rosa Lönneborg
For family and friends.
List of publications


II. **Lönneborg R** & Brzezinski P. Factors that influence the response of the LysR-type transcriptional regulators to aromatic compounds. *BMC Biochem* 2011, **12**:49

III. **Lönneborg R**, Varga E & Brzezinski P. Directed evolution of the transcriptional regulator DntR-isolation of mutants with improved DNT-response. *Submitted manuscript*


# These authors have contributed equally to the publication

Additional publications


Abstract

The primary aim of the work presented in this thesis was to change the inducer specificity of the DntR protein in order to improve the response to DNT. The long-term goal is to use this protein in a biosensor for DNT, a signature compound for detection of the explosive TNT. Another aspect of this work was to understand the mechanisms of inducer binding and how the binding of an inducer molecule changes the DntR structure into a state that triggers transcriptional activation.

In the included papers, the inducer specificity of wt DntR has been investigated under different conditions. The functional effects of specific mutations have also been investigated, in some cases in combination with structure determination using X-ray crystallography. In addition, structural data offering insights into the details of inducer binding and conformational changes upon inducer binding are presented and discussed in terms of mechanisms for transcriptional activation by DntR. Furthermore, a directed evolution strategy was employed in order to find variants of DntR with improved response to DNT. A variant with a large improvement in the DNT response was isolated and characterized. In optimized growth conditions, this DntR variant had a nearly 10-fold increase in fluorescence in response to DNT compared to wt DntR. Specific substitutions found in this DntR variant are suggested to be important for changing the inducer response.
Contents

Introduction ............................................................................................................................... 9
  1.1 Life and evolution ............................................................................................................ 9
  1.2 Transcriptional regulation in bacteria ........................................................................ 11
  1.3 The evolution of pathways for degradation of xenobiotic compounds ... 13
    1.3.1 Characteristics of nitrotoluenes .......................................................................... 15
    1.3.2 Degradation of 2,4-dinitrotoluene (DNT) ......................................................... 15
  1.4 Biosensors .................................................................................................................... 17
    1.4.1 Biosensors for detection of environmental pollutants ........................................ 19
    1.4.2 Biosensors for detection of DNT ........................................................................ 21

2. The LysR family and DntR ......................................................................................... 23
  2.1 General characteristics of the LysR family ................................................................. 23
    2.1.1 Functional characteristics .................................................................................... 23
    2.1.2 Structural characteristics .................................................................................... 26
  2.2 The LysR-type transcriptional regulator DntR ............................................................ 30
    2.2.1 The full-length DntR structure ............................................................................ 32

3 Protein engineering ....................................................................................................... 34
  3.1 Rational design: site-directed mutagenesis ................................................................. 34
  3.2 Directed evolution ....................................................................................................... 35
    3.2.1 Generation of library diversity .............................................................................. 35
    3.2.2 Error-prone PCR ................................................................................................... 36
    3.2.3 Oligo-nucleotide directed mutagenesis (site-saturation mutagenesis) .................... 38
    3.2.4 Recombinatorial techniques .................................................................................. 38
  3.3 Screening of cell-based libraries .................................................................................... 39
    3.3.1 Flow cytometry and FACS ................................................................................... 40

4. Present investigations ..................................................................................................... 41
  4.1 Development of a whole-cell prototype biosensor ...................................................... 41
    4.1.1 E. coli as a host organism for the DntR-based biosensor .................................... 41
    4.1.2 Using gfp as a reporter gene ............................................................................... 43
    4.1.3 The influence of growth conditions and gene arrangement on the inducer response .................................................................................................................. 44
  4.2 Structural investigations of DntR ................................................................................ 45
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNT</td>
<td>2,4-dinitrotoluene</td>
</tr>
<tr>
<td>DBD</td>
<td>DNA-binding domain</td>
</tr>
<tr>
<td>IBC</td>
<td>Inducer-binding cavity</td>
</tr>
<tr>
<td>IBD</td>
<td>Inducer-binding domain</td>
</tr>
<tr>
<td>FACS</td>
<td>Fluorescence-activated cell sorting</td>
</tr>
<tr>
<td>LTTR</td>
<td>LysR-type transcriptional regulator</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>RD</td>
<td>Regulatory domain</td>
</tr>
</tbody>
</table>
Introduction

1.1 Life and evolution

Since the first appearance of life, organisms live in constant interaction with their environment. The first living organisms were likely to be far more simple than even the least complex life-form existing today. Noteworthy, there are some features that distinguish all living organisms from dead matter; compartmentalization, metabolism, reproduction and passage of hereditary information to coming generations.

An organism must be separated from its environment (compartmentalization) in order to maintain balance in the chemical processes necessary for life to continue. However, a living being must also be able to transport compounds from and to its surroundings, and respond to changes in the environment (Figure 1). Compounds imported from the surroundings are metabolized in order to obtain energy and building blocks for the more complex macromolecules that enable the organism to perform work, maintain cellular structures and to reproduce.

Figure 1. Principle of a living cell. A cell is separated from its environment, but is also in constant interaction with its surroundings. Molecules (cube) are taken up and processed in order to obtain energy and building blocks (rectangles). Rest products (small squares) have to be transported out of the cell. DNA contains genetic information that is used for construction of proteins, which are involved in various functions of the cell.
When the organism has reproduced, the hereditary information (DNA in modern life forms) is passed on to the offspring, thereby passing on particular features of the parent organism. In most cases, the offspring will not be an exact copy, due to the occurrence of mutations and recombination that result in diversity among the offspring. Many changes are deleterious, meaning that they would make the organism less fit for survival in that particular environment. A few changes would be beneficial and therefore give this particular offspring an advantage. Some changes would be neutral, giving neither advantage nor disadvantage to the carrier of this change. The individuals carrying beneficial changes would have a survival advantage and produce more offspring than the individuals carrying deleterious changes. In this way, the beneficial changes will be passed on to new generations, while deleterious changes will die out. This process is called natural selection, the central concept of evolution. For the neutral changes there is neither negative nor positive selection pressure. Over time, these neutral changes can introduce a large variation within a population without giving a selective advantage-a process called genetic drift. In another environment or in combination with additional mutations, a previously neutral change may become beneficial or deleterious, and thus be subjected to natural selection.

All organisms living today are presumed to share a common ancestor among the earliest living organisms, referred to as the Last Universal Common Ancestor [1]. From this common ancestor, life spread and diversified into various organisms that are divided into three domains: Archaea, Bacteria and Eukarya. The domains of Archaea and Bacteria are commonly referred to as prokaryotes since both domains include single-cellular organisms that lack nuclei. Cells that have a nuclear membrane surrounding the chromosomal DNA are classified as eukaryotes. Eukarya is the domain where multicellular organisms such as plants, animals and fungi are found. In Eukarya, many organisms increase the genetic diversity in the offspring by sexual recombination, while diversity introduced by mutations plays a less dominant role. Among prokaryotes there is no typical sexual recombination, and random mutations was first thought to be the only mechanism for introducing diversity among bacteria. However, since the first evidence of lateral gene transfer in prokaryotes was discovered [2] there is growing evidence that it plays a major role in the evolution of bacterial genomes [3,4]. In lateral gene transfer, genetic material is transferred between any two organisms that may come from even very distantly related species, as opposed to the vertical gene transfer that occurs between one generation and the next generation within the same species. Lateral gene transfer can be mediated via three main mechanisms; conjugation, bacteriophage-mediated transduction and transformation [5].

Once genetic material is acquired, the genes need to be expressed properly and, with proper timing, regulated with the rest of the genome. The acquisition of genes from other species via lateral gene transfer, higher mutation rates and fast
generation times enable bacteria to adapt with amazing speed to changes in the environment. Indeed, prokaryotes have adapted to almost any environment found on earth and they are found in such different ecological niches as glaciers, hot sulphuric springs, the acidic stomach of humans, the deep ocean and the dry environment of deserts. Sequencing of DNA isolated from natural habitats has revealed larger diversity in the microbial community than previously thought to exist [6,7]. In recent years, humans have introduced drastic changes to the environment and reduced many ecological niches. This has led to the extinction of many species, but thereby also introduced new possible niches where evolution of new metabolic pathways has been made possible.

1.2 Transcriptional regulation in bacteria

All organisms from the three domains of life share the same basic principle for replicating, transcribing and translating the hereditary information (Figure 2). In order to express gene products only when necessary, thus avoiding spending energy on producing proteins with no use at the time, genetic regulation is needed. The proportion of genes devoted to genetic regulation tends to increase with the genome size and the complexity of the environment in which the organism lives [3]. In general, the genetic regulation is less complex in prokaryotes than in eukaryotes. In Bacteria the main genetic regulation occurs at the level of transcription initiation [8], although regulation also occurs at other levels.

Figure 2. The central dogma of molecular biology. The DNA is transcribed into RNA, which is translated into a protein. In order to transfer genetic material when the cell divides, the DNA is replicated.
In Bacteria, all transcription into mRNA is dependent on one RNA polymerase that has to bind to a specific DNA sequence, the promoter. The promoter is situated upstream of a gene or an operon in order to start transcription. The interaction of RNA polymerase with different promoters is dependent both on different σ-factors, and on specific transcription factors [9]. A σ-factor binds to the core RNA polymerase and is required for the RNA polymerase to recognize promoter sequences [8]. Different σ-factors recognize different sets of promoters. In *Escherichia coli*, there is a main σ-factor, σ^70^, which enables the RNA polymerase to recognize most promoters, but *E. coli* also has six other σ-factors that are associated with the core RNA polymerase in response to different stress situations. In these cases the other types of σ-factors bind to other specific subsets of promoters upstream of genes needed in these stress situations [10]. Different bacterial species have different sets of σ-factors, due to the need to recognize different environmental stimuli.

The role of the transcriptional regulators is to sense specific signals from the surroundings and regulate the expression of genes so that the organism can adapt to changes in the environment. In many cases, transcriptional regulators require the presence of a specific σ-factor for proper transcriptional activation [10-12]. This interplay between σ-factors and transcriptional regulators enables integration of specific signals (detected by the transcriptional regulators) with more general physiological signals (mediated by the type of σ-factor). In other cases, the general physiological state of the cell affects the activity of transcriptional regulators in still unknown ways. Additional layers of complexity in regulation are given by the possibility for other factors (such as anti-σ-factors, anti-anti-σ-factors and proteases) to modulate the activity or turn-over of transcriptional regulators and σ-factors [8].

The signal recognized by a transcriptional regulator can be transduced either directly (one-component signal transduction system) or indirectly (two-component signal system). In the one-component system the transcription regulators usually binds a small signal molecule that has been taken up by the cell, and then regulate transcription of genes directly. In the case of indirect signal transduction, a signal is detected by a receptor protein on the surface of the cell, and then further mediated via phosphotransfer relay [13] or other intra-cellular signal systems such as cAMP or c-di-GMP dependent signalling [3] to the transcriptional regulator. Transcriptional regulators usually bind to one or several specific promoters where they can modify the expression of genes downstream of a promoter in several ways. They can interact with the promoter and prevent the RNA polymerase to position itself for transcriptional initiation thereby acting as a repressor. A repressor may also bind to distant DNA sequences to create looping of the core promoter, and thereby make the promoter inaccessible for the RNA polymerase. A third way for a repressor to prevent transcription is to modulate another transcriptional regulator that acts as an activator [9]. An activator is a type of transcription factor that facilitates transcriptional initiation
either by direct interaction with RNA polymerase or interaction with a σ-factor or another transcriptional regulator. Many transcriptional regulators can function both as repressors and activators, depending on the presence/absence of different signal molecules.

In general, a transcriptional regulator does not regulate the expression of only one gene, but a cluster of genes called an operon. The operon is usually transcribed into a single mRNA that is spliced into several products, translated into proteins involved in the same pathway [14]. There are also global regulators, transcriptional regulators that regulate several operons and genes (these clustering of operons with coordinated regulation are called regulons) [15].

Typically, transcriptional regulators consist of a small-molecule binding domain and a DNA-binding domain, although variations exist. The majority of the DNA-binding domains found among prokaryotes are variations on the helix-turn helix domain [16], but there are also less common DNA-binding domains such as the Zn-finger, helix-loop-helix, β-sheet antiparallel DNA-binding domains and RNA-binding-like domains [15]. However, these types of domains appear to be constrained to specific genomes and are not universally spread. There are about 75 identified families of transcription factors among prokaryotes, where the LysR family is the largest family found so far [15,17] (this family is described in section 2.1).

1.3 The evolution of pathways for degradation of xenobiotic compounds

During the 19th century, humans learnt how to synthesize organic compounds. Since then, an enormous number of chemical compounds that previously did not exist in Nature have been produced by man. Chemical compounds that cannot be metabolically produced or used by an organism are referred to as xenobiotics. Such compounds are in general toxic to most organisms. The easiest way for an organism to handle toxic compounds is to evade exposure, either by moving away from them or pump them out of the cell(s). However, new biochemical pathways where these xenobiotic compounds can be completely degraded have emerged in microbial organisms living in polluted environments [18]. Degradation enables the use of the xenobiotic compound as a source of energy and/or useful building blocks instead of using energy to get rid of the xenobiotic. Evolving pathways for degradation therefore provides a large adaptive advantage for an organism in a polluted environment. However, this ability poses a larger challenge than just getting rid of the xenobiotic, since a whole
pathway consisting of several enzymes has to be recruited and the enzymes have to be adapted to a new task and also be properly regulated.

The major biochemical pathways that exist today have evolved during billions of years and are today impressively efficient with well-regulated enzymes displaying high specificity. However, biochemical pathways for degradation of man-made xenobiotics are in general mediated by poorly regulated, low specificity enzymes [19]. From an evolutionary perspective, the existence of these far from perfect pathways is not surprising, considering the very short evolutionary period of time that these xenobiotics have been present in the environment. Genes encoding different “precursor” enzymes could be recruited from various organisms via lateral gene transfer and be patched together into new pathways, and these already existing enzymes may obtain the ability to accept new substrates after only few point mutations [18]. This so called patchwork theory predicts that the result is a pathway consisting of enzymes with the ability to use both the old and new substrates with an often low efficiency, resulting in broad specificity “promiscuous enzymes” [20]. Sometimes, already existing enzymes may be able to use xenobiotics as a substrate, although with low efficiency [18]. To achieve high specificity for a new substrate, more mutations have to accumulate, the new substrate must be available and use of the substrate should give the organism an adaptive advantage.

Enzymes recruited for a new function in a pathway for xenobiotic degradation are generally not subjected to efficient regulation. Often they are first expressed at a low constitutive level [21]. However, in a complex mixture of different growth substrates, efficient regulation of different catabolic pathways enables an organism to use the best growth substrate available at the time. Also, as more species emerge that can explore the same ecological niche (i.e. use the same substrate as a nutrient), more efficient regulation becomes a necessity [22]. Thus, the transcription factors and also promoters involved in regulation of a degradative pathway appear to evolve with similar mechanisms as described above for enzymes; at first there is a reduced repression in the absence of an inducer (resulting in low constitutive expression of regulated enzymes) accompanied by a relaxation of inducer specificity [23], and only later a higher specificity for a metabolically relevant inducer/effector evolves [21]. Usually, efficient regulation also involves coupling and integrating the response to a specific ligand to signals associated with the general physiological state of the cell (stress signals, metabolites from basic metabolism etc.) [22].

There is evidence for a high degree of lateral gene transfer in microbial communities living in heavily polluted environments [18]. Here, genes or part of genes can be transferred on different mobile genetic elements between different species in the microbiota. Also, different species may perform different parts of the degradation of a chemical compound [24], such that a complete pathway for degradation does not necessarily evolve independently in one organism.
study of these degradation processes is of importance for the understanding on how pollutants can be naturally removed from the environment (bioremediation). In addition, these emerging pathways for degradation of xenobiotics provide many interesting opportunities for evolutionary biologists to study details of intermediate steps in an ongoing evolution.

1.3.1 Characteristics of nitrotoluenes

From heavily polluted environments, Fungi and Bacteria have been isolated that are capable of degrading toxic and persistent xenobiotics such as monoaromatic hydrocarbons (e.g toluene, benzene) [25-27], polyaromatic hydrocarbons [28], chlorinated aromatic compounds [20,29] and nitro-aromatic compounds [30-34]. Very few nitro-aromatic compounds existed in nature before the era of synthetic chemistry. One group of such man-made nitro-aromatics are the nitrotoluenes [35].

Nitrotoluene consists of an aromatic carbon ring with a methyl group (toluene) with 1 to 3 nitro-substituents. Trinitrotoluene (TNT), first synthesized in 1863, has been used extensively as an explosive since 1902 and continues to be one of the most used explosives today. Discharges of TNT from munitions factories and usage of TNT in large amounts as an explosive have resulted in contamination of soil and groundwater both with TNT and derivatives thereof. Due to the toxicity and to the explosive nature of TNT, there is an increasing demand for quick and easy detection of TNT in various environments. One of the main components derived from TNT production and release is 2,4-dinitrotoluene (DNT) [36,37]. DNT is present in higher amounts than TNT itself in the vapour phase and aqueous solutions derived from commercial TNT samples [36]. It is not normally produced in other natural processes (with the exception of very low amounts formed by the photochemical reaction of toluene with nitrogen oxides [38]. This makes DNT a suitable signature compound/molecular marker for the presence of TNT [36,37]. DNT is also produced as a starting material used in the synthesis of polyurethane foam [34], but due to its toxicity, modern formulations try to avoid the usage of DNT.

1.3.2 Degradation of 2,4-dinitrotoluene (DNT)

DNT is a relatively stable molecule, but the electron-withdrawing nitro-groups can be reduced into the corresponding amino-group with relative ease. Reduction of the nitro-group is also the most reported biological degradation occurring in bacteria and fungi isolated from TNT/DNT contaminated soil and water [35,39], but some species are also capable of complete removal of the nitro-group, using it as a nitrogen-source [35]. Such reactions are often quite slow.
and most often result in dead-end products with no complete mineralization. It is also possible under some conditions for a reduced amino-group to be oxidized back into a nitro-group so that DNT is reformed (redox-cycling) [30,35,40]. The conversion between nitro- and amino-aromatics occurs via highly reactive nitroso- and hydroxylamino- intermediates that can cause adduct formation with biomolecules, which is presumably the major cause for the toxicity of DNT. The reductive degradation of nitrotoluenes can be carried out by non-specific nitro-reductases that can be found in intestinal microflora and mammalian organs [39]. It is unclear whether the reductive degradation products possess similar toxicity as DNT itself or are less toxic [41-44].

**Figure 3.** The oxidative degradation pathway of 2,4-dinitrotoluene (2,4-DNT) in *Burkholderia* sp. strain DNT. 4M5NC: 4-methyl-5-nitrocatechol. 2H5MQ: 2-hydroxy-5-methyl-quinone. 2,4,5THT: 2,4,5-trihydroxytoluene. DHMOHA: 2,4-dihydroxy-5-methyl-6-oxo-2,4-hexadienoic acid. DntR regulates the expression of the operon containing dntA dntB and dntC encoding the enzymes DNT-dioxygenase, MNC-monooxygenase and benzenetriol-oxygenase. Figure adapted from [45].

In contrast to the non-specific reduction of DNT, oxidation and complete mineralization of DNT (via cleavage of the aromatic ring) make it possible to use DNT as a growth substrate. Two isolated *Burkholderia* strains have been reported as capable of oxidative degradation of DNT. In these strains, DNT could be used as the sole source of both carbon and nitrogen resulting in complete mineralization of DNT [34,46]. The genetic elements involved in the oxidative pathway of DNT have been mapped in both these strains, and the degradation appears to occur via very similar pathways, where the genes encoding the enzymes needed for the first steps of the pathway are nearly identical between the two strains (Figure 3). The sequence and the genetic organization differs more for the genes encoding enzymes involved in the last steps [46]. The first enzymes in the pathways were regulated by the same transcription factor named DntR in both species (differing only in one amino acid substitution, S192T between the two strains). The “wild type” (wt) DntR used in the work described in this thesis originates from *B. cepacia* R34, with a serine at position 192.

In general, examination of the organization of genes and the function of enzymes needed in the catabolic pathways of nitro-aromatic compounds reveals evidence of ongoing evolution such as remnants of lateral gene transfer, scattered organization of the genes encoding the enzymes, inefficient regulation and poorly adapted enzymes with low catalytic specificity [19].
This is indeed true for the DNT degradation pathways as well. DNT dioxygenase encoded by the genes dntAa-d, regulated by DntR (Figure 3), has very high similarity to naphthalene-dioxygenase and has been demonstrated to use naphthalene as a substrate more efficiently than DNT [47,48]. In addition, several Pseudomonas strains have the ability to degrade naphthalene, using a catabolic pathway regulated by NahR (60 % sequence identity to DntR), where salicylate is an intermediary metabolite and also functions as an inducer for NahR [49]. There are also remnants of an inactive salicylate hydroxylase gene in the region encoding the DNT pathway genes, which has 99 % sequence identity with a gene in the naphthalene degradation pathway regulated by NagR (99 % sequence identity to DntR) [46]. This indicates a recent divergence of the gene cluster involved in the first steps of DNT degradation from a gene cluster involving genes in a naphthalene degrading pathway. In the later steps of the DNT-degrading pathways, the elements involved differ more between the two isolated Burkholderia species and seem to have been recruited from different sources [46].

Emerging pathways for degradation of other nitro-aromatic compounds have also been found, where two pathways regulated by NtdR deserve to be mentioned [50]. These two pathways were found in two different strains isolated for growth on 2-nitrotoluene and nitrobenzene respectively. NtdR that regulates the expression of the first enzymes in both these pathways shows very high sequence similarity to DntR and is also induced by salicylate (see section 4.3).

1.4 Biosensors

A biosensor is a device that uses biological material for detection of a specific chemical substance (the analyte) where the biological material interacts with the analyte and produces a measurable signal either by itself or in combination with some physical/chemical device. In general, the performance of a biosensor is highly dependent on the specificity and the sensitivity of the interaction of the biological material with the analyte. Other important factors for robust and reproducible analysis are the stability of the biological material used for detection and the possibility to perform a reliable calibration. The biological material in question could be a protein such as an enzyme, regulatory protein or antibody, but also a nucleic acid, bacteriophage or a whole cell.
Figure 4. Example of a whole-cell biosensor. Microscopic picture of E. coli cells that express the green fluorescent protein (GFP) that is under control of the P_{DNT} promoter. When an inducing compound is present, the transcriptional regulator DntR activates the transcription of gfp and the cells become fluorescent.

Whole-cell biosensors have the advantage of being produced easily in large amounts to a very low cost, due to the rapid reproduction of most cells. They also measure the bioavailable concentration of a compound, which is often an advantage. Disadvantages are that transport in and out of the cell and biodegradation of the analyte can lead to fluctuations in the output signal (lower robustness). The sensitivity is in general also lower than for a cell-free system. An isolated biomolecule has the potential of giving a more sensitive and reliable signal. The output signal could also be faster if binding of the analyte can be measured in real time. However, the preparation of a cell-free system is in general more complicated and requires more expensive equipment. The isolation of the biomaterial in question may also be expensive and complicated, depending on the source and purification method required.

Biosensors have today been developed for detection of a large number of different analytes. Due to problems such as difficulties with robustness and low specificity, many prototype biosensors are still in the need of further development before they can compete with standard analytical methods. However, there are an increasing number of commercially available biosensors today. In clinical applications enzyme/antibody-based assays are used today for detection of compounds such as glucose, metabolites or different hormones in samples from patients, and in food industry biosensors are tested for product control and detection of contaminants [51]. Also, biosensors are applied for detection of illicit drugs and explosives. Another growing area where biosensors are applied is for detection of various pollutants in the environment [52].
1.4.1 Biosensors for detection of environmental pollutants

An increasing number of biosensors has been reported that monitor the presence of environmental pollutants. In these applications, biosensors often have an advantage over traditional analytical techniques due to their ability to detect bioavailable concentrations of the analyte. The actual concentration might be higher, but if a fraction of the chemical is bound tightly to particles or other matter, this fraction could be unavailable for uptake by/interaction with biological systems. This fraction may then be irrelevant when considering factors such as biological reactivity or toxicity. Thus, for many applications the bioavailable concentration is more relevant to measure.

Other conventional sensors often rely on an extraction step, where the amount of the analyte that is extracted may differ considerably from the bioavailable concentration. In general, conventional analytical methods are often able to detect analytes with higher sensitivity, but are more expensive and time-consuming than most biosensor systems [51]. Biosensors could preferably be used in applications where screening for toxicity/biological reactivity of large sample-volumes in situ is required, where the lower cost and rapid monitoring of the bioavailable concentration without any extraction steps are considerable advantages when compared to conventional analytical methods. There are also biosensors that monitor more general features, such as general toxicity (where the stress response to DNA-damaging agents is measured) in environmental samples [53]. An example where biosensors would be the preferred choice due to the low cost is field-monitoring of environmental pollutants in large areas of soil or water.

Both whole-cell and cell-free biosensors have been reported, where most of the biosensors are based either on an enzymatic assay or on a whole-cell system. The whole-cell biosensors often rely on a transcription factor capable of binding the analyte, causing an increase in gene expression of a reporter gene fused to a promoter that the transcription factor acts upon [54]. The reporter gene would typically give expression of a protein that would produce an easily detected output signal such as light or colour (fluorescence). Examples of common reporter genes are gfp [55], lacZ and lux [56]. There are also examples of whole-cell systems relying on other components than a transcription factor, for example, the use of a membrane-bound receptor protein [57]. Cell-free biosensors, where the biomaterial detecting the analyte is attached to a surface, have also been reported [58-60].

Often the transcription factor component of a typical whole-cell biosensor system has originally been isolated from sites contaminated with the pollutant in question, such as the toluene/xylene-sensing XylR [61], toluene-sensing TbuT [54], mercury-sensing MerR [62] and naphthalene (salicylate)-sensing NahR.
In most cases, the sensitivity and specificity of the original transcription factors have been low. Attempts to alter the sensing capability using different mutagenesis approaches have been tried in several cases, such as for NahR [64], XyIR [23], TetR [65] and the phenol-sensing DmpR [66]. The sensitivity could also be improved by simple technical adaptations as in the case of NahR, where detection could be improved 10-fold by measuring naphthalene in the gas-phase instead of the aqueous phase [67]. The magnitude of the response can also be amplified by constructing multi-component regulatory cascades with more than one transcription regulator responding to the same inducer [68]. Another approach to increase the level of response is by optimizing the promoter sequence to which the transcription factor binds [69].

Another complication with using whole-cell biosensors is the complexity of the bacterial cell; the response to the presence of a signal molecule is also dependent on other global regulatory signals regarding the metabolic state of the cell. This integration with other signals leads to variations in the output signal that also depends on other factors such as growth conditions and metabolic stressors [21] (see section 1.2), which weakens the robustness of the response. However, optimization and calibration could be used to reduce the effect of these variations. Also, design of the signal detection circuit can be done so that stability and robustness are improved [70].

The sensitivity of different biosensors varies within a wide range; detection limits have been reported from 2 nM [62] to 0.55 mM [58]. In most cases, an improvement in the sensitivity is needed in order to use the biosensor in practical applications. Also, the specificity is usually too broad for many applications of the reported biosensors, which may result in false positive responses thereby limiting practical applications of the biosensors. However, in some cases a broad specificity may be desired. For example, a group of compounds, such as PCBs or phenols [71], that possess similar chemical properties (such as persistence or toxicity) also often occur together as a mixture of pollutants. Detection of many components simultaneously can in these cases be an advantage, provided that the specificity includes the members of such a group of chemical compounds and exclude other types of compounds. In addition, transcription factors that are induced by an environmental pollutant can also be used as improved regulators for pathways of degradation in bacterial strains applied in bioremediation. In this application, high specificity is not as important as long as the regulation is functional, i.e. the transcription factor should upregulate the expression of genes needed for the degradation of the pollutants to be removed.
1.4.2 Biosensors for detection of DNT

In recent years, several studies have reported prototype biosensors for 2,4-DNT [23,58,59,72,73]. In the work by Galvao et al. [23], the bacterial transcription factor XylR, that originally was found to respond to toluene, was modified in order to respond to DNT. A mutant library was constructed and mutants responding to DNT were isolated. The isolated mutants all displayed a broad specificity and a low sensitivity, responding to a lowest reported concentration of 125 µM DNT in solution while still retaining a high response to toluene and some other toluene-derivatives [23]. This XylR variant was in a later study integrated into the chromosome of \textit{P. putida} KT2440 and used for detection of added DNT in soil samples, with luciferase as a reporter [70]. In this study the lowest amount of DNT that gave a detected response was 3-4 mg spread over a soil sample.

In a study using a different strategy, engineered rat olfactory receptors were expressed on the surface of yeast cells, and a screen was performed to identify DNT-responsive olfactory receptors, using GFP as a reporter for activation. Several clones that responded to a DNT concentration of 25 µM DNT were identified, but no test of the specificity was reported [73]. By expressing the sensor protein on the surface of the cell, the problem of transporting the compound into the cell is circumvented, which shows potential to lower the threshold of detection. However, yeast cells are not suitable as a host for release in an environment outside the laboratory. The only report so far of a whole-cell biosensor that provided a robust signal in soil was for the low-sensitivity XylR mutant in \textit{P. putida} [70].

Another strategy for increasing the sensitivity is to use a cell-free system, with either the analyte or the sensor attached to a surface. In the study of Nagatomo et al. [59], a cell-free system was used where polyclonal antibodies detecting DNT were allowed to bind to a DNT-analog attached to an ethylene glycol-coated gold surface. DNT could be detected using a Surface Plasmon Resonance (SPR) technique with very high sensitivity, detecting DNT in the concentration range of 5-500 nM with a linear response. The selectivity of the assay was tested with 8 analogous compounds, where a higher response to 2,4-dinitrophenyl-glycine than for 2,4-DNT itself was seen, and a response in the same range as for DNT was seen for 2-amino-4,6-dinitrotoluene and 1,3-dinitrobenzene [59]. A system based on surface-attached oligopeptides detecting DNT using a Quartz Crystal Microbalance (QCM) technique has also been developed [58]. Oligopeptides previously selected with phage-display to bind DNT and TNT, respectively [72] were in the study of Cerruti et al. attached to a
poly(ethylene-co-glycidyl methacrylate)-coated gold-silicon surface and DNT detected at a concentration of 550 µM (although the best TNT-binding peptide had a dissociation constant of 71 nM as determined using ITC). The selectivity of the DNT-binding peptide was only tested with TNT.

Detection of DNT currently relies on standard analytical chemical techniques with an extraction/adsorption step of the analyte followed by analysis using gas chromatography or HPLC [36,37]. The lowest level of detection reported for 2,4-DNT by GC analysis is in the low nM range when extracting the analyte from soil [37]. There are also examples of prototype chemical sensors for detection of DNT, such as the sensor developed by Naddo et al. [74], where a fluorescent polymer-film was allowed to interact with DNT- or TNT-saturated vapour. Adsorption of DNT or TNT to the polymer-film caused fluorescence quenching, which yielded a measurable signal [74]. However, the specificity of this detection method is unclear, since the π-π stacking interactions described in Naddo et al [74] most likely could be provided by numerous other aromatic compounds. Also, the sensitivity, i.e. detection limit was not reported.
2. The LysR family and DntR

2.1 General characteristics of the LysR family

2.1.1 Functional characteristics

The LysR family is the largest known family of transcriptional regulators among prokaryotes [15,17]. The members of this family are 300-330 amino acids long, and have an N-terminal DNA-binding domain and a C-terminal effector-binding domain [75]. The two domains are connected via a flexible linker region. The DNA-binding domain consists of a winged helix-turn-helix motif (wHTH) with a high degree of conservation at the sequence level, while the effector-binding domain shows large sequence variation.

The effector is usually a small molecule that can act as an inducer for transcriptional activation of genes downstream of the promoter that the LysR-type transcriptional regulator (LTTR) acts upon. However, there are exceptions where small molecules instead function as anti-inducers that lead to repression of genes downstream of the promoter [76,77], or where the switch between activation and repression appears to occur without binding of a ligand [78]. The kind of effector molecule to be recognized varies greatly, as do the genes and operons that the LTTRs regulate. LTTRs are involved in regulation of pathways for a diversity of functions such as metabolism, cell division, virulence, motility, toxin production, nitrogen fixation, oxidative stress response and degradation of xenobiotics to name a few [17]. There are both LTTRs that function as local regulators of specific operons and some LTTRs that function as global regulators [79].

In general, the LTTR is always bound to the promoter that it regulates, both in the absence and presence of an inducer (Figure 5). For most LTTRs, the promoter also functions as a dual promoter, regulating the genes downstream of the promoter in one reading direction while at the same time auto-repressing its
own gene that is transcribed in the opposite reading direction. It has been proposed that this organization of the LTTR and its target gene would facilitate lateral gene transfer, thereby at least partly explaining the high abundance of LTTRs among prokaryotes [15].

Details of the promoter DNA-LTTR interactions have been studied using DNaseI footprinting and similar techniques for several LTTRs such as ClcR [80], CatR [80,81], TcbR [82], CrgA [83], CbbR [84], OccR [85], BenM [86] and NahR [49]. The results from these studies showed that there was a similar protection pattern of the promoter for most LTTRs, although some variations exist. Typically, an LTTR binds to 50-60 bp in the promoter region [75], but some LTTRs can also bind up to 80 bp of DNA. Some LTTRs such as NahR [87] are known to recognize more than one promoter region.

The second and the third helix of the wHTH are responsible for binding to the major groove of DNA and there is also evidence that some conserved arginine residues in the “wing” are important for interactions with the minor groove of DNA [88,89]. The promoter region that LTTRs bind to in general contains a regulatory binding site (RBS) and an activation binding site (ABS) (Figure 5). In general the interaction between the protein and the RBS is stronger than between the protein and the ABS [75,80]. There is a consensus sequence, T-N_{11}-A within the RBS in the promoter region called the LysR motif that is generally recognized by LTTRs, likely by contact with conserved arginine residues in the wing [88]. No conserved sequence appears to be found for the lower affinity ABS site [17], and in most cases the binding to the ABS site is shifted upon inducer binding to a second ABS site (Figure 5A).

There are also examples where there is no binding to the ABS in the absence of inducer, where some LTTRs, such as BenM and CatR, instead bind to an internal binding site (IBS) within the controlled operon [81,86] (Figure 5B). Usually the ABS site closely proceeds or overlaps with the -35 position that is recognized by the σ-factor of RNA polymerase. Whereas most LTTRs appear to be tetramers in their biologically functional forms [17,75] there are also examples of LTTRs that appear to form dimers or higher-order oligomers [90] in their functional form. The type of interactions with the promoter DNA is expected to be different depending on the multimerization state, due to the different number of DNA-binding sites present in the functional protein.
Figure 5. Schematic picture of transcriptional activation by tetrameric LTTRs. In A) a ClcR type mechanism and in B) a CatR type mechanism.

Binding of an inducer molecule to the inducer-binding domain (IBD) is proposed to induce a conformational change that propagates to the DNA-binding domain (DBD) and changes the interaction with the promoter DNA. DNAseI footprinting / gel mobility shift assays have shown for many LTTRs (NahR [49], CysB [91], OxyR [92], OccR [93], CbnR [94] and CbbR [84]) that binding of an inducer molecule affects the structure of the protein-promoter complex such that the bending angle of the promoter DNA is relaxed. When this bending angle is relaxed, transcriptional activation of the genes downstream the promoter can occur. For CysB, it has been shown that a part of the DNA-binding domain (between α2 and α3) contains an activating region that interacts with the C-terminal domain of the α-subunit (α-CTD) of RNA polymerase [95], responsible for activating transcription of the genes downstream of the promoter. Also for CatR, ClcR and NahR, interactions with the α-CTD of RNA polymerase have been demonstrated [80,96].

Models for how the LTTR binds to DNA and how the presence of an inducer affects DNA binding and induces transcriptional activation have been proposed and are shown schematically for ClcR-type interaction with the promoter (Figure 5A) and for CatR-type interaction (Figure 5B). For BenM, interactions with the promoter is similar to that of CatR, but it is complicated by the fact that two different compounds (benzoate and cis-cis-muconate) can function as inducers in a synergistic fashion [86]. The same RBS and ABS sites are protected from cleavage in DNAseI footprinting with BenM in the presence of benzoate or cis-cis-muconate, as well as when both inducers are present. However, hypersensitive bands disappeared when both inducers were present, indicating that the bending angle was further relaxed compared to when only one inducer was present.
2.1.2 Structural characteristics

In recent years several crystal structures of members of the LysR family have been determined. In the past it was difficult to obtain enough pure, soluble protein for crystallization of full-length LTTRs [97]. Nevertheless, over the last years several structures of full-length LTTRs have been solved; for ArgP [98], TsaR [99], CrgA, [89], CbnR [100] and DntR [101]. There are also two structures representing putative LysR-type proteins with unknown function. Both these proteins originates from *P. aeruginosa*; one deposited in the PDB with the code 3FVZ as the outcome of a structural genomics study [99] and the other deposited with the PDB code 2ESN [102]. Recently, another full-length structure of BenM was published, but the packing of dimers in the crystal lattice resulted in an asymmetric arrangement so that no tetramers representing a biologically functional tetramer were found [102]. Whereas CbnR [100], DntR [101], BenM [102] and TsaR [99] are tetrameric in solution, CrgA [89] is claimed to be octameric or a double octamer in its functional form. The oligomerization state of ArgP [98] remains unclear. ArgP was crystallized as two different types of dimers, one with interactions between the DBDs and the other with interactions between the IBDs. ArgP was dimeric in solution as shown by gel filtration, but was suggested to form tetramers (a dimer of the two different types of dimers) when bound to DNA, at least in the presence of inducers, based on observations of the interaction of ArgP with promoter DNA.

The first full-length structure to be solved, that of CbnR [100], revealed that the monomer was organized into a DBD consisting of the first 57 N-terminal amino acids, which is connected to the IBD (a.a.91-294) with a linker region (a.a. 58-90) (Figure 6A, Figure 7B). The IBD fold is similar to the fold of periplasmic substrate-binding proteins and could be further subdivided into two subdomains; regulatory domain 1 (RD1) and regulatory domain 2 (RD2), connected by a hinge. Both RD1 and RD2 display a Rossman fold topology. Between RD1 and RD2, a cleft is formed that is proposed to be a general site for effector binding (Figure 6A). No effector molecule was found in this cleft, and CbnR was therefore proposed to represent a repressor-state conformation of the protein. CbnR formed a tetramer in the crystal structure, which was shown using other methods to correspond to the biologically active form. The four identical subunits (A, B, P, Q) in the tetramer are arranged as a dimer of dimers, where each dimer is composed of two monomers in a head-to-tail arrangement (Figure 6B).
**Figure 6: A schematic illustration of the structure of LTTRs.** In A): the structure of one monomer with one DNA-binding domain (DBD) in red connected to the inducer-binding domain (IBD) in green/blue via a linker region (yellow). The position of the inducer binding cavity (IBC) is marked with a red circle. B) The homo-tetramer is formed from interactions between four monomers. The monomers are identical in sequence, but adopt different conformations in the tetramer (one extended and one compact). The arrangement of the monomers in the tetramer varies greatly. In some cases, the main interactions takes place between the central IBDs (and between the DBDs to form a DBD pair), and in some cases interactions in the tetramers occurs mainly in the linker and DBD regions.

A comparison of the other full-length structures of LTTRs [89,98,99,101] reveal that they all have a very similar fold of the monomer as described above for CbnR (Figure 7B). However, the tetramers adopt significantly different conformations such that the positions of the IBDs both relative to each other and relative to the DBDs are different (Figure 7A-D). In all structures, one monomer in a head-to-tail dimer adopts a compact form, while the other monomer adopts a more extended form. In the monomer with compact conformation, the linker between IBD and DBD folds back over the surface of the IBD, while this linker protrudes away from this surface in the extended conformation (Figure 6 and Figure 7). The angle between the IBD, the linker and DBD varies both for the compact form and the extended form for the different LTTR structures, making the positioning of the DBD pairs quite different [102]. For all tetrameric full-length structures, the linkers brings two DBDs (one from an extended and one from a compact monomer) together in a pair that could bind to major grooves in either RBS or ABS in the promoter region [17] (Figure 5).
Figure 7. Structures of full-length tetrameric LTTRs. Each monomer is shown in a different colour (dark and light green, dark and light blue). To the left: Side-view. In all structures, the DBD from each green and blue monomer forms two apposed DBD pairs, but the position of the DBD pairs is different for A-D. To the right: Central view (rotated 45° to the right), where differences in interactions in the central part of the tetramer can be observed. A) The model of full-length DntR. The tetramer adopts a compact conformation that is stabilized by interactions between the RD2 subdomains of opposing monomers. B) The compact CbnR tetramer with interactions between contact helices in RD2 domains of light blue and light green monomers (but not in dark green and blue monomers). C) 3FZV tetramer with interactions both between RD2 subdomains and RD1 subdomain, but also a central cavity. D) TsaR tetramer in an extended conformation with a large central cavity, and interactions mainly between the linkers.
Also, a central cavity is seen between the four IBDs in TsaR and the structure with the PDB code 3FZV. This central cavity is almost absent in DntR, and in CbnR this space is occupied by two helices from opposing monomers (Figure 7, central views). This central cavity reflects different interaction surfaces between the IBDs in the different protein structures. The two helices seen for CbnR, referred to as “contact helices” are presumed to be important for the structural rearrangement that occurs when an inducer binds to the protein [99]. Structural alignment of the pairs of DBDs shows that the distance between the DBD pairs is 50 Å in the case of TsaR, 62 Å in the case of CbnR and 80 Å in the case of 3FVZ [99]. The different distances between the pairs of DBDs are expected to reflect differences in the distances between the positions of RBS and ABS-binding sites within the promoter element either by basepair distance in the different promoters or by bending DNA differently. The bending would also be different depending on whether the LTTR represents an activator state (that promotes transcriptional activation) or a repressor state (that prevents transcriptional activation).

Monferrer et al. proposed that the compact CbnR structure represents the repressor state of CbnR [100], while both the TsaR structures without the inducing compound para-toluene-sulfonate (TSA) and with TSA present are proposed to represent the activator state [99]. It is known that CbnR binds to the same binding sites in both the presence and absence of its inducer cis-cis-muconate but binding of the inducer results in a relaxation of the bending angle in the promoter DNA [94]. However, the interaction of TsaR with its promoter is yet not known, which makes it difficult to discuss its mechanism of transcriptional activation in detail. Therefore, the model proposed by Monferrer et al. [99] remains to be proven.

So far, it has not been possible to crystallize a full-length LTTR both in the repressor conformation with empty IBC and in an activator conformation with ligand bound. Soaking of crystals of TsaR with TSA gave no change in the structure compared to the conditions without TSA, suggesting that the protein was locked in one state in the crystal [99]. However, soaking of crystals of truncated BenM (lacking the first N-terminal 80 amino acids) with the known inducers cis-cis-muconate (CCM) and/or benzoate resulted in structures of the truncated IBD with a slightly different conformation than the crystals with no ligand bound [103]. As expected, CCM was bound to BenM in the cavity between RD1 and RD2 regarded as a general IBC of LTTRs. Surprisingly, an unexpected benzoate molecule was found at a second binding site in the hydrophobic core of RD1 in one of the monomers in a dimer. In the structure derived from crystals soaked with benzoate, benzoate was found at both these binding sites. The structural changes within the monomer are relatively small between the conformation with no ligands, only one ligand in the primary IBC and one ligand in each binding pocket. However, in a model of the full-length tetramer based on the CbnR structure, the small conformational change in the monomer
upon ligand binding leads to large changes in the interactions in the tetramer. This model suggests that binding of ligands would move RD1 and RD2 slightly closer together, which would result in a compressed tetramer, resulting in a large movement of the DBD pairs closer to each other. This compressed tetramer would then represent the activator state of the protein. This is consistent with the previous DNaseI footprinting pattern observed for BenM [86] (see section 2.1.1). However, this model is opposed to the model for TsaR by Monferrer et al [99] which suggests that the compressed form of the tetramer represents the repressor state, while a more expanded conformation is responsible for transcriptional activation.

So far, the truncated DntR structures presented in paper IV are the first reported where the native inducer is co-crystallized with the LTTR protein. In these structures, binding of ligands appear to induce a similar movement of RD1 towards RD2, resulting in a more expanded tetramer as in the model by Monferrer et al [99], but the conformational change (when two salicylate molecules are bound/monomer) is much larger (see section 4.4). If the structure without salicylate bound represents the repressor state and the structure with two salicylate molecules bound represents the activator state, then DntR acts using a similar mechanism as proposed for TsaR by Monferrer et al. [99].

2.2 The LysR-type transcriptional regulator DntR

One of the members of the LysR family is DntR, the protein on which this work is focused. The dntR gene was isolated from two different Burkholderia species that were capable of complete oxidative degradation of DNT [34,46]. Because DntR was found to regulate the expression of genes needed for the first steps of the DNT-degradation, it was assumed to be activated by DNT or a direct metabolite thereof. The results from an early study of wild-type DntR expressed in P. putida as a prototype biosensor for detection of DNT indicated that there was a slight response to DNT, but the response to salicylate was much larger [104]. This response profile is not surprising, considering the high sequence similarity of DntR with homologous LTTRs that respond to salicylate, such as NahR, NtdR and NagR (Figure 8) and the short evolutionary time-span for a DNT-response to evolve.

DntR is a tetrameric LTTR, which has been shown to bind to a 120 bp DNA fragment containing the promoter region (P_{DNT}) between the transcriptional starts of the dntR gene and the dntA gene. The affinity for the promoter was slightly higher in the presence of the inducer salicylate [101]. Salicylate is not an intermediate metabolite in the DNT-degrading pathway, but probably still functions as inducer due to the high sequence similarity between the proteins involved in the first steps of DNT degradation and the genes involved in naph-
thalene degradation (see section 1.3.1). Indeed, only two amino-acid substitutions in the DBD differ between DntR and NagR, a transcription factor which is induced by salicylate that was found to regulate a naphthalene degradation pathway in *Ralstonia* sp. strain U2 [105]. NahR, another regulator of naphthalene-degradation is also induced by salicylate and has 60% sequence homology with DntR. DntR also shows high similarity with NtdR (7 amino acids differ) that regulates the degradation of 2-nitrotoluene and nitrobenzene in two different bacterial strains. NtdR has broader inducer specificity than DntR, still responding to salicylate but also to several nitro-aromatic compounds [50] (paper II). The sequence similarity to other LTTRs reported to regulate degradation of aromatic compounds is relatively low (Figure 8).

**Figure 8.** An unrooted tree derived from ClustalW alignment of amino-acid sequences retrieved from GenBank for some selected LTTRs. With the exception of LysR that regulates the synthesis of proteins needed for lysine biosynthesis, they all regulate genes encoding proteins needed for degradation of aromatic compounds. LTTRs where salicylate functions as an inducer are marked with an orange circle. A blue diamond indicates that some other aromatic compound(s) functions as an inducer(s). A green triangle marks the regulators where cis-cis-muconate or close derivatives thereof function as inducers. For LysR, diaminopimelate functions as an inducer.
2.2.1 The full-length DntR structure

Structures were previously solved for the full-length DntR protein with isothiocyanate and acetate found in the proposed inducer-binding cavity (IBC) [101]. The electron density of the DBD and the linker between the DBD and IBD was too poor to obtain good resolution data of these regions, but the structure of the IBD was solved at 2.3 Å for the DntR structure with thiocyanate and 2.6 Å, with acetate. The IBD in the full-length structure shares the same overall fold with other LTTRs, with two Rossmann fold subdomains, RD1 and RD2. In a third crystallization condition, the overall electron density was lower, but the electron density of the DBD was higher. Altogether, data obtained from all three crystal conditions were sufficient to propose a model for the full-length DntR (Figure 9). The positions of the linker helix and two of the helices in the HTH domain were determined in the model, but not refined. Also, all three crystal conditions showed packing of the monomers into a homotetramer, consisting of a dimer of dimers, very similar to that of CbnR [100].

As in CbnR, two monomers of DntR interact in a head-to-tail fashion with strong interactions between the IBDs while the DBDs of these monomers point in opposite directions. However there are also interactions between the DBDs from distant monomers from each head-to-tail dimer (Figure 9). The DBDs from two such monomers form a DBD pair with a distance between the recognition helices of 31 Å (i.e. about the same as for CbnR, 30Å). This distance corresponds well with the distance between two major grooves in DNA (30-33 Å) and also the T-N11-A LysR-motif. Also like in CbnR, one monomer in each dimer has an extended structure while the other has a more compact structure. However, the position of the IBDs compared to the DBDs in both monomers that form a DBD pair is markedly different between CbnR and the model of DntR. For example, the positions of the two DBD pairs are much further apart in DntR compared to CbnR (Figure 7 A-B and 9). This implies that also the distance between the RBS and the ABS elements in the promoter is different, either due to a different basepair distance between the RBS and ABS or due to a difference in the bending angle.

Both salicylate and DNT were modelled into the proposed IBC in the structure with acetate, and on the basis of this model, mutations that would improve DNT-binding were proposed. The effect of some of these mutations were investigated and the data are presented in papers I and II (see "Present studies" below). More recently, structures were solved for a truncated variant of DntR (consisting of residues 90-301) with only the IBD, where salicylate was co-crystallized with the protein at two crystal conditions and the IBC was without any other ligands than water in another crystal condition (paper IV and further described in "Present studies").
Figure 9. Model of a homo-tetramer of full-length DntR with DNA. The model is based on the structure of the IBD with thiocyanate in the IBC, while the structure and position of the DBDs are obtained from a third crystallization condition (resolution 3Å). DNA oligos are positioned so that the second and third helices are placed in the major groove. Each monomer is coloured in a different colour, DNA is light-grey. A pair of DBDs from two different monomer is expected to interact with the RBS site and the opposite pair of DBDs with the ABS site of the promoter.
3 Protein engineering

Recombinant DNA technology [106] and the polymerase chain reaction (PCR) method [107] made it possible to isolate genes from any natural source, amplify the genetic material and to express the proteins encoded in another host organism. Different methods for altering the properties of a certain protein soon emerged, creating the field of protein engineering. However, it is not trivial to introduce specific properties into a protein. A protein is a highly complex and very large molecule, where any change in a certain amino-acid residue might give rise to unpredicted effects on the folding and flexibility/stability properties of the protein [108]. The strategy of relying upon available information in order to improve the properties of a protein by making specific amino-acid substitutions is often referred to as rational design.

In contrast, another strategy called directed evolution has emerged, where changes are more or less randomly introduced into the sequence, creating a large number of protein variants. By using an efficient selection method for finding the few protein variants with improved properties, protein variants with the desired property can be isolated.

3.1 Rational design: site-directed mutagenesis

Rational design relies on the ability to introduce specific mutations into a gene encoding the protein that is to be altered. Today, several different PCR-based techniques exist that introduce mutations with the help of primers with mismatched bases. The challenge of rational design is to predict what amino-acids to substitute in order to obtain the desired alterations of the protein properties, for instance the affinity of a ligand. A good prediction relies on having adequate information about the protein in question, preferably with high resolution structural data as a basis. Computer-assisted simulations may be of great help in order to predict protein-ligand interactions after in silico mutagenesis [57].

Site-directed mutagenesis has been applied in many cases to improve catalytic properties of enzymes [108,109]. However, the effect of even a single mutation can give rise to unexpected effects. Frequently, unpredicted changes in protein
structure and stability occur. Also, the structural data obtained from X-ray crystallography gives a static picture of a protein, and assumptions made from structural data may be misleading due to lack of knowledge about the dynamic interactions of the protein with its ligand.

In the case of DntR and other signal-transducing proteins, where binding of a ligand is coupled to a global conformational change that affects the whole protein structure, the effect of mutating residues could be especially difficult to predict. Residues involved in ligand binding are often also involved in propagating the conformational change to the rest of the protein, giving rise to unpredictable effects on the protein stability and transcriptional activation [110]. Also, substitutions in parts of the protein that are not involved in ligand binding might strengthen or weaken the response (measured as transcriptional activity) when the ligand binds. In cases where binding of a ligand is accompanied by a large global conformational change, amino-acid substitutions in segments involved in propagating the structural change may alter the functional properties of the protein [111]. Such structure-function relations are not easily predicted. In these cases, a directed evolution strategy to change properties of the protein might provide better results.

3.2 Directed evolution

In contrast to rational design, a directed evolution strategy does not necessarily require any detailed knowledge about the function and structure of the protein that is to be modified. Directed evolution-based methods instead rely on the principle of evolutionary selection to find variants with desired properties from a collection of clones (referred to as a library) with more or less randomly introduced variations that create sequence diversity. From this library of clones with different sequences, a few variants may by chance carry changes that would improve a desired property. Challenges in using a directed-evolution strategy lay both in creating a library with good quality (i.e. sequence diversity where mutations that improve desired properties are likely to occur) and in identifying an appropriate selection method that would make it possible to identify variants with improved properties.

3.2.1 Generation of library diversity

Variation can be introduced using many different methods. Chemical mutagenesis was one of the earliest methods used to introduce mutations into bacterial strains [112], where addition of different mutagenic chemicals led to introduction of mutations into the whole genome of an organism. This approach
could be useful when looking for changes in global properties such as altered metabolic properties, but this method is less suited for changing specific properties of one protein, where mutations preferably should be introduced into the gene encoding this protein (or regulation of this gene in some cases). Another way to introduce mutations on a global genome scale is to use mutator strains that have a higher mutation rate than other strains [113].

However, when changing a specific gene sequence (which is usually the case when changing the properties of a protein) there are still many different alternatives. Several techniques exist to introduce variations into a specific gene sequence, such as variants of error-prone PCR and oligo-nucleotide-directed randomization that introduce mutations with higher or lower degree of randomness. Variation can also be obtained with various recombination methods, where already existing homologs of a gene are mixed (Figure 10).

![Figure 10. Different approaches to generate library diversity. A) Random mutagenesis. Mutations are introduced randomly into the gene segment with methods such as ep-PCR or variations thereof. B) Oligo-nucleotide directed mutagenesis. Only specific nucleotides are selected for randomization. C) Recombination. Pre-existing variants (homologues) of a gene segment are recombined into chimeric variants.](image)

### 3.2.2 Error-prone PCR

In error-prone PCR (epPCR) a gene or a fragment of a gene is amplified using mutagenic conditions during the PCR reaction, resulting in random mutagenesis (Figure 10A). Low fidelity polymerases (i.e. lacking good proof-reading) could be used, such as Taq DNA polymerase. However, even the low fidelity of Taq DNA polymerase (1 error in about 9 000 bases [114]) would not generate
enough diversity for the construction of a gene library. The error rate of a polymerase could be enhanced by using conditions such as increased amounts of Mg\(^{2+}\), adding small amounts of Mn\(^{2+}\) or having unequal amounts of dNTPs in the reaction mixture [115,116]. Using an enzyme such as Taq DNA polymerase under mutagenic conditions would introduce a bias in the construction of the library due to the inherent propensity of this enzyme for making transitions with a higher probability than transversions. In a transition, structurally similar bases are exchanged (i.e. A->G, T->C) whereas in a transversion a pyrimidine base such as A or G can be replaced by a purine such as T or C. Thus, some codons are less likely to be obtained in a library created using Taq DNA polymerase.

In recent years, DNA polymerases have become commercially available that are tailor-made for having very low fidelity and to make transversions with as high probability as transitions. By using such an enzyme, the traditional bias of epPCR can be avoided (as in paper III). However, in all variants of epPCR, stop-codons are introduced that would result in expression of truncated protein variants. Also, due to the natural degeneracy of the genetic code, there is typically a “codon bias” where some amino acids are more likely to be obtained than others. There are 61 codons encoding 20 amino acids and three stop codons. Consequently, some amino acids are encoded by more than one codon. For example, leucine is encoded by 6 codons, while tryptophan is encoded by only one codon. Also, a single mutation in a codon cannot give rise to all of the other 19 amino acids. To convert a particular codon into some of the other codons, two or three mutations are required [117]. On average, a single random mutation in a gene can give rise to codons encoding 5.7 other amino acids [109].

However, one can argue that the bias arising from the natural degeneracy of the genetic code has been subjected to natural selection during billions of years, and it is amazingly stable with only minor modifications in some of all living organisms today [118]. The codons are therefore likely to be optimized for achieving a good balance between conservation of protein structure/stability and plasticity in function, so this bias might in fact be an advantage when constructing a diverse gene library, since it helps to reduce the number of mutations that would cause loss of function.

There is also another type of bias introduced by the amplification process in PCR. A mutation that is introduced early in the amplification process is amplified along with additional rounds in the PCR and thus, it would be over-represented in the final pool of PCR products [117]. This bias can be reduced by performing several amplification reactions in separate tubes with fewer reaction cycles.
A total randomization of a gene encoding 100 amino acids would generate a library size of $20^{100}$ (>2400) variants of this gene. A library of this size would be impossible to cover during selection. With fewer codons accessible as when only single point mutations are introduced, the theoretical maximum size of the library is reduced, but still $20^{5.7}$ variants is usually too large to handle for practical applications. In practice, libraries are often limited in size by bottlenecks in cloning such as ligation and transformation and also by the possibility of screening the complete library. Therefore, it is often desirable to construct a less randomized library but with the aim of introducing changes that more likely would introduce the desired properties.

3.2.3 Oligo-nucleotide directed mutagenesis (site-saturation mutagenesis)

Oligo-nucleotide directed mutagenesis is a semi-rational approach, where random mutagenesis is applied only at certain selected codons of a gene (Figure 10B). This approach requires an accurate prediction of which sites to mutate in order to change a given property. By using this approach the total number of variants is reduced but the likelihood of finding variants with improved properties could be higher. Technically, total randomization of one codon could be achieved by amplifying the gene with PCR using internal primers that are synthesized with randomly inserted bases (NNN) at the position of this codon. It is also possible to restrict the mixture of bases in the last position (NNT/C, NNG/C or NNT/C/G) in order to avoid variants that would introduce stop codons. There are also techniques for synthesis of oligonucleotides using defined sets of triplet nucleotides that code for specific amino acids in order to avoid bias in the resulting libraries [119,120].

3.2.4 Recombinatorial techniques

Another way to create sequence diversity is to use a technique based on recombination instead of introduction of mutations. This approach is analogous to the homologous recombination that occurs naturally in Eukaryotic meiosis. In a recombination process, homologous/similar sequences are fragmentized and then the fragments are joined in new combinations (Figure 10C). Often, homologous genes from different species have been used for recombination, but sequence diversity may also be created using various mutagenesis techniques prior to recombination (as in paper III). A common method used to create recombinatorial libraries is to use DNAse I digestion to create small fragments of genes with a subsequent PCR step, where the small fragments are used as primers that successively reconstruct a mosaic full-length gene sequence consisting of fragments from several different parental sequences. This method requires
careful optimization in order to get DNAse I-digested fragments of optimal length. Another method that can be used is the Staggered Extension Process (StEP) [121]. In StEP, the template sequences are first subjected to a PCR process, where the annealing step is very short and at a non-optimal temperature, resulting in copying only fragments of the template sequences. When cycles are repeated, fragments anneal to different templates. After enough rounds this procedure results in a full-length mosaic sequence. To get enough copies of the recombined sequence, a standard PCR reaction with normal PCR conditions can be performed. StEP is a very fast method to create sequence diversity for a recombinatorial library and can be performed in less than a day. The degree of variation in the resulting recombinatorial library depends both on how similar the original gene variants are and on the size of the fragments that are obtained in the first part of the process.

3.3 Screening of cell-based libraries

Once a library of mutants is obtained, the major challenge when using a directed evolution strategy is to find an appropriate selection method. In order to find mutants with improved properties, a large library size is often advantageous and screening of variants to find a few among a vast majority could resemble the old problem of finding a needle in a haystack. Usually screening is performed on the basis of distinguishing an improved phenotype that is linked to a changed genotype. Problems may arise when selecting for a phenotype that can be linked to more than one genotype (which is quite often the case since the function and expression of a protein are affected by many processes in a cell). Careful design of the selection method with the aim of reducing the number of false positives is therefore important.

Colony screening using an assay that would distinguish positive clones from other clones on the basis of survival advantage or colour change was the first used method for screening and is still in use. However, this method is very time-consuming and laborious, and today more efficient methods for screening are available. There are several methods for surface display of proteins such as phage display [122], ribosome display [123] and yeast display [124] that have been proven successful in many cases for high-throughput screening of libraries for improved binding affinity. However, these display techniques are difficult to apply when screening for other properties than binding, such as enzyme activity or transcriptional activity. In these cases, linking activity to formation of a fluorescent product [108] or expression of a fluorescent protein [125] enables the use of fluorescence-activated cell-sorting (FACS) for high throughput screening (see paper III).
3.3.1 Flow cytometry and FACS

Flow cytometry is used to measure the properties of single cells or particles as they pass by a light source (usually a laser). The light emitted from the laser is scattered by each particle, and the scattering depends on the particular properties of the cell or particle, making it possible to distinguish different types of cells/particles on the basis of their light scattering and light emitting properties. Labelling with or expression of fluorescent molecules is often used, since fluorescence enables very sensitive detection [126]. Different biomolecules can be fluorescently labelled in order to obtain specific information about the distribution of for instance a certain receptor within a cell population. A powerful extension of flow cytometry is the possibility to sort cells, so called fluorescence-activated cell-sorting (FACS) (Figure 11). The possibility to sort cells can be used to isolate cells with unusual properties from a large population, which could be used for high-throughput screening of recombinant protein libraries [127,128] (paper III). Several rounds of sorting are generally necessary for enrichment of very rare variants.

Figure 11. Principle of Fluorescence Activated Cell Sorting (FACS). Cells (or other particles) are suspended in fluid, and this cell-suspension is focused in a narrow stream of liquid. The flow is adjusted in order to get separation of individual cells. The stream of cells is subjected to vibrations at high frequency so that the stream breaks into very small drops (droplets). Ideally, there is only one cell/droplet. Before the stream breaks into droplets, the stream of cells passes a laser that analyzes the passing cells. At the point where the stream breaks up into droplets, an electrical charging ring can apply an electrical charge to the “last hanging drop”. If a cell has the desired properties for being sorted, the electrical charging ring will apply a charge to the droplet that contains that cell. The charged droplet will then be deflected by the voltage plates so that it can be physically sorted out into some type of container (test tube, 96 well plate, petri dish etc.).
4. Present investigations

4.1 Development of a whole-cell prototype biosensor

In order to study binding of different inducers to DntR, a gfp reporter gene was used to obtain a measurable signal of DntR-mediated transcriptional activation in whole cells. The gfp gene was placed under the control of the P<sub>DNT</sub> promoter to which DntR binds. In this construct, addition of an inducing compound to E. coli DH5α cells harbouring the dntR and P<sub>DNT-gfp</sub> components leads to DntR-mediated transcription and expression of the green fluorescent protein (GFP). GFP fluorescence can be detected on the whole-cell level using methods such as flow cytometry. In the work described in paper I, a two plasmid system was used to evaluate the response to various potential inducers (Figure 12A) in E. coli. This system was compared to another plasmid arrangement for expression of wt DntR in both E. coli and P. putida, and the response to salicylate was comparable (4-5 fold induction), while the response to DNT was negligible in both organisms and both plasmid systems. The two-plasmid system expressed in E. coli was therefore selected to analyze the response to the various potential inducers for different mutants of DntR. These mutants carried substitutions lining a proposed inducer binding cavity and had been suggested based on an earlier study [101].

Later on, this two-plasmid system was compared to a one plasmid system (Figure 12) that increased the sensitivity of the inducer response (paper II). This one-plasmid system was then used also in the studies described in papers III and IV.

4.1.1 E. coli as a host organism for the DntR-based biosensor

E. coli is one of the most extensively studied organisms today. The bacterium was early selected as a model organism since it is easy to grow, DNA manipulations can be done easily, and most strains are non-pathogenic. Non-pathogenic wild-type strains are part of the normal flora of the gut, while most laboratory
strains have lost the ability to thrive in the intestine. Most standard molecular biology techniques used today are optimized for E. coli, which greatly facilitates cloning and expression of recombinant proteins when using E. coli compared to other bacteria. For the work presented in paper III, where construction of a large number of clones (~10^7) was required, the high transformation efficiency that was achieved would most likely have been impossible in any other host organism. Another advantage of using E. coli rather than the original DNT-degrading strains during the development of the biosensor is that there are no specific degradation pathways for DNT in E. coli. As a consequence, DNT is not removed at the same rate as in the DNT-degrading Burkholderia strains and the ability for the biosensor to respond to DNT persists for a longer period of time.

However, TNT has been reported to be degraded in some E. coli strains by yet unidentified enzymes with nitro-reductase activity. NemA, NfsA and NfsR have been suggested to be the enzymes involved in TNT degradation [129,130]. Similar non-specific reduction of DNT is expected to occur in E. coli. An indication of such a process is the formation of some compound(s) with absorption maximum at ~360 nm in E. coli cultures exclusively after addition of DNT (paper II). Also, characterization of a specific mutant that was isolated in the work described in paper III suggests that the improved response to DNT found in this mutant is in fact not a direct response to DNT, but rather a response to a metabolite of DNT formed several hours after addition of DNT (unpublished data). However, the DNT-response does not correlate in time with the formation of the compound(s) with absorption ~360 nm. The DNT-response in the mutant with increased response is elevated with high cell density where various metabolic factors may be modulating the response. For instance, the salicylate-induced response and the DNT-induced response could be dependent on different σ-factors, as in the case of mutants of the XylS protein [131].

It is also likely that there are differences in the intracellular concentration of DNT between E. coli DH5α and the original DNT-degrading strains. In E. coli, the cells appear to pump DNT or metabolites thereof out of the cell as indicated by the observation of the yellow colouring of the growth media after DNT-addition. In contrast, in the original DNT-degrading B. cepacia R34, a putative ABC transport system that was suggested to be involved in transport of DNT into the cell was found on the same plasmid as other genes involved in the DNT-degrading pathway [46].

Also in the case of NtdR, the ability to recognize nitro-aromatic compounds was lost when NtdR and its promoter were transferred to P. putida or E. coli from the original nitro-aromatic degrading strains [50]. This loss of response may, at least partly be explained by differences in transport mechanisms in the different strains. However, several other explanations for the strain-specific recognition of nitro-aromatic compounds in the case of NtdR are also possible. There could
be differences in global regulatory networks involved in catabolite repression [132]. Also, a higher degree of non-specific nitro-aromatic reduction in the original strains, in combination with that this reduced product is the true inducer of NtdR could give higher response in these strains.

For practical use of a biosensor in contaminated soil, *E. coli* is not a suitable host, since the *E. coli* strains used would not survive in soil outside of the laboratory. In addition, a plasmid-based system is not stable in the host and usually contains antibiotic resistance genes that should not be released into the environment. However, once a DntR-based transcription factor with good enough specificity for DNT is obtained in *E. coli*, it could be transferred to another host strain. A genetic system for transferring of bioreporter genes to *P. putida* KT2440 that ensures stable integration into the chromosome and exclusion of antibiotic markers has been developed previously [70].

4.1.2 Using *gfp* as a reporter gene

The reporter gene *gfp* [133] that is used in the prototype biosensor has several advantages. The fluorescent protein GFP requires no co-factors for activity, no substrate has to be added, and the fluorescent signal enables very sensitive detection even at relatively low expression levels. GFP is also a very stable protein, which may be an advantage or a disadvantage depending on the application [55]. However, other commonly used reporters such as different luciferases and β-lactamase require the presence of co-factors and/or substrates, but have been reported to give higher sensitivity and faster detection compared to *gfp* [69,134,135]. The output fluorescence signal of *gfp* is directly correlated to the level of GFP protein expression, thus giving a linear correlation between transcriptional activation and the measured signal (disregarding contribution from additional regulation of mRNA concentration and translation rate). When using luciferase that generates light or β-lactamase that generates colour in an enzymatic reaction, an amplification of the signal occurs, giving higher sensitivity and faster detection. As a consequence, the lifetime of the signal is also shorter with luciferase and β-lactamase, so in applications where the response is followed over a long time, GFP is the preferred choice. When following whole-cell fluorescence, GFP could accumulate in cells that have stopped dividing, giving a significant signal increase even after the cells have entered the stationary phase while still expressing the reporter.
4.1.3 The influence of growth conditions and gene arrangement on the inducer response

In the work described in paper II, the importance of the gene arrangement for the sensitivity of the system and also the influence of different growth conditions were investigated. The sensitivity of the inducer response was improved several-fold when the dntR gene was placed under control of the PDNT promoter in a one plasmid system as compared to the two plasmid system used in paper I (Figure 12). The presence of a His6-tag at the C-terminal of DntR also reduced the inducer response. In addition, the background fluorescence in the absence of inducer was lowered when changing growth media from a rich medium (Luria-Bertani) to a modified minimal medium (paper II). This could be due to the presence of an inducing compound in the rich growth media or due to other global metabolic signals that affected transcription and/or translation.

Figure 12. Plasmid arrangement used in the prototype biosensor. A) The “two plasmid system” where the dntR gene is expressed from a T5/lac promoter. B) The “one plasmid system” where the dntR gene is autorepressed and expressed from its own promoter (PDNT). Bla is the β-lactamase gene, giving ampicillin resistance and neo gives neomycin/kanamycin resistance.

The fold of induction when salicylate was added did not differ significantly for the LTTRs expressed in the one-plasmid system when the cells were grown in the different growth media. However, for mutants of DntR with broader specificity, the response to DNT (and also benzoate and 2-nitrobenzoate) increased in the minimal medium compared to LB (paper II). Analysis of the 5.3p3c19 mutant that was isolated in the work described in paper III displays the same pattern (but here also the salicylate response is lower in LB). Further characterisation of this mutant indicates that the response to DNT is delayed in time compared to the salicylate response. This result suggests that the response to DNT addition is indirect and originates from metabolite(s) formed in E. coli (unpublished data). The formation of degradation products is expected to be at least to some extent regulated in response to general metabolic conditions such as nitrogen limitation and might therefore increase in the minimal medium.
4.2 Structural investigations of DntR

The previously determined structure of wt DntR gave a good overview of the overall protein fold [101]. The structural investigations in this work have instead focused on determining where an inducer binds and how the binding of an inducer affects the structure of DntR. In the work described in paper I, the structure of two mutants with mutations lining a proposed inducer-binding cavity (IBC) were solved (Figure 13) with thiocyanate found in the previously proposed inducer-binding pocket [101]. By substituting amino-acid residues lining this cavity, a change in inducer specificity was observed in some cases. In other cases, the ability to respond to any of the potential inducers tested was lost. These results strongly suggested that the previously proposed position of this IBC was correct.

In the work described in paper IV, additional structures of the IBD of wt DntR were solved, with and without salicylate found in the previously identified inducer-binding pocket. However, from some of the crystal forms with salicylate, a second salicylate moiety was also found in a cleft at the end of RD1 where the linker to the N-terminal domain is positioned (see Figure 16). This observation, together with other data suggests that there could be two inducer-binding sites per monomer of DntR.

4.2.1 The primary inducer binding pocket

In the first structure of the full-length DntR, a putative inducer-binding cavity (IBC) was identified between RD1 and RD2 in the IBD [101]. In the work described in paper I, the position of this IBC was validated by mutations of residues lining the IBC. In one of the two mutants the substitution of tyrosine at position 110 with a serine and the substitution of phenylalanine 111 with a valine were proposed to enlarge the IBC in order to improve DNT binding. However, the structure of this Y110S/F111V double mutant revealed that these substitutions instead resulted in a decreased volume of the IBC due to a different rotamer of the side-chain of leucine 151, and movement of both F167, R248 and H169 compared to the previous wt structure (Figure 13, left). This mutant responded equally well to salicylate and benzoate, while wt DntR was more specific for salicylate. In the structure of the wt IBD with salicylate (paper IV) it is seen that that H169 provides a hydrogen bond to the hydroxy-group of salicylate, and that benzoate in the same position would have no interactions with any residue in the hinge region. In the Y110/F111V mutant, however, the smaller size of the cavity (due to flipping of L151) would allow the carboxy-group to form interactions with H169 as well as with H206 and L151, thus providing the same possibility for benzoate as salicylate to make contact with RD1, RD2 and the hinge region upon binding.
In contrast, the structure of a mutant with the phenylalanine replaced by a leucine (F111L) revealed an enlargement of the hydrophobic part of the IBC (Figure 13, right). Also here, the side-chain of H169 moves significantly and H206 flips to a rotamer that occupies a small cavity close to the IBC that is unoccupied in the wt structure. On the other hand, Y110, F167 and R248 do not move significantly. The enlarged cavity of the F111L mutant would provide room for the methyl-group of DNT that does not fit into the hydrophobic part in the earlier modelling studies of DNT-binding in the IBC of wt DntR [101]. However, there was no apparent DNT response in vivo for this mutant, and a very low response to salicylate and benzoate (Table 1). This could be explained by later findings demonstrating the importance of the hydrogen bonding of the inducer to H169 and H206 for stabilizing a conformation that is presumably required for transcriptional activation (paper III and paper IV).

Co-crystallisation of the IBD of DntR with salicylate confirmed that salicylate binds to the IBC (paper IV). However, the position of the hydroxyl-group of salicylate is mirrored compared to the position in the previous model. Residues from both RD1 and RD2 and the hinge region participate in the binding of salicylate, resulting in the closed conformation of the IBD. The aromatic ring of salicylate fits well into the hydrophobic part of this pocket lined with the side chains of I106, G107, Y100, F111, G152, F167, R248 and I273 (Figure 14). A number of hydrophilic interactions closer to the entrance of the cavity provide specific binding to the carboxyl and hydroxyl group of salicylate. Direct hydrogen bonds are formed between the hydroxyl group of salicylate and the side chain of histidine 169. A hydrogen bond is also formed between the carboxyl-group of salicylate and H206 in RD1. There are also several water-mediated hydrogen bonds between salicylate and main chain atoms of T104, L151 and L153. In addition, interactions are seen between the aromatic ring of salicylate and the aromatic side-chains of F167, Y110 and F111 in the hydrophobic inner part of the cavity (Figure 14).
Interestingly, the details of salicylate-binding in the primary IBC is affected by the presence of a salicylate molecule found in an additional binding site (Figure 15). In the case where the additional salicylate is present, two water molecules are excluded from the primary IBC, and a hydrogen bond is formed between the carbonyl oxygen of leucine 153 in RD1 and the amide hydrogen of glycine 205 in RD2 (Figure 15 B). This interaction stabilizes a closed conformation where the β8-α8 and β9-α9 loops in RD2 moves closer to the β4-α6 loop of RD1, rotating the two subdomains and moving them closer to each other. The structure with only one salicylate molecule bound on the other hand, is nearly identical to the open conformation with no salicylate present (Figure 16).
4.2.2 A putative second inducer-binding site

In the work described in paper IV, a construct with the first 90 N-terminal residues (ΔN90-DntR) and a construct with the first 80 N-terminal residues (ΔN80-DntR) deleted were co-crystallized with salicylate. Some crystal forms with the ΔN90-DntR construct were found to have two salicylate moieties / IBD monomer. In these crystal forms the IBD had a much larger conformational change compared to the previous full-length structure (rmsd Cα 1.58-2.21, rotational angle 14-21°) than in the structures where only one salicylate was found in the primary IBC (rmsd Cα 0.87-1.06, no rotational angle) (Figure 16).
The salicylate found in this secondary IBC has less specific interactions and had two different binding modes in crystal forms I and III (Figure 15A), where the carboxy-group is either hydrogen-bonded to S95 or via a water to D284. The aromatic ring of salicylate is situated in the same cleft as F90 in the model of full-length DntR [101] and as the aromatic ring of W90 in TsaR [102] in the monomers with a compact conformation. In the extended conformation, this residue is in a different conformation, leaving this cleft unoccupied. It is possible that a ligand in this cleft helps stabilizing the linker in a proper position for the transcriptionally active form of the protein. Substitution of F90 by an alanine resulted in a mutant with no response at all to salicylate, which suggests that the interaction between the linker and IBD at this cleft is crucial for transmitting the conformational change of the IBDs to the DBDs to activate transcription. Also, substitution of S95 to an alanine and R97 to a cysteine resulted in mutants with lower response to salicylate (paper IV). However, this secondary inducer binding-cleft was unoccupied in the ΔN80-DntR construct that was co-crystallized with salicylate. Although some data support the suggestion that the second binding site is functionally relevant, further investigations are needed to confirm this proposal.

*Figure 16. Comparison of the ΔN90-DntR construct with and without salicylate present.* In A) the structures without (grey) and with one (pink) salicylate bound, respectively. In B) the structure without salicylate (grey) and two salicylate bound (blue) are shown. No significant domain movement is observed in the truncated monomer of DntR in A) when there is only a salicylate found in site 1 (primary IBC). In contrast, a large movement of RD2 (bottom) relative to RD1 (top) is observed when a salicylate is found also at site 2 (the putative second inducer-binding site). This movement results in a more closed conformation of the monomer.
4.3 Alterations of the inducer specificity of DntR

The inducer specificity was tested both for wt DntR and a large number of DntR variants with a gfp gene under the control of the P\textsubscript{DNT} promoter as a reporter for transcriptional activity. DntR was grown in the absence or presence of a selected group of potential inducers with chemical similarities to salicylate or DNT. The effect of these different potential inducers was then compared for wt DntR and different mutants under different growth conditions and different plasmid arrangements (paper I-III).

4.3.1 Changes of inducer specificity observed in different mutants created by rational design

In the work described in paper I, a number of mutants were constructed on the basis of earlier suggestions of structural alterations that would improve the binding of DNT into the putative IBC [101]. Several of these structural variants did not show response to any of the potential inducers, suggesting that these mutations resulted in an inability of DntR to function as a transcriptional activator. Some mutants however, displayed less specificity for salicylate with an improved response to benzoate compared to wt DntR (Table 1). One of these mutants, F111L/H169V, also displayed further broadening of the inducer response and also responded to DNT and 4-nitrobenzoate during these conditions.

<table>
<thead>
<tr>
<th>DntR variant</th>
<th>SAL</th>
<th>DNT</th>
<th>2-NT</th>
<th>4-NT</th>
<th>4-HT</th>
<th>2-NB</th>
<th>4-NB</th>
<th>BEN</th>
</tr>
</thead>
<tbody>
<tr>
<td>wt</td>
<td>5.0</td>
<td>1.0</td>
<td>1.1</td>
<td>0.7</td>
<td>0.9</td>
<td>1.1</td>
<td>1.0</td>
<td>1.3</td>
</tr>
<tr>
<td>Y110S/F111V</td>
<td>3.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.3</td>
<td>1.3</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>H169V</td>
<td>3.2</td>
<td>1.1</td>
<td>1.0</td>
<td>0.8</td>
<td>1.1</td>
<td>1.1</td>
<td>1.3</td>
<td></td>
</tr>
<tr>
<td>F111L</td>
<td>1.7</td>
<td>0.9</td>
<td>1.0</td>
<td>0.8</td>
<td>1.2</td>
<td>1.1</td>
<td>1.5</td>
<td></td>
</tr>
<tr>
<td>F111L/H169V</td>
<td>4.2</td>
<td>1.6</td>
<td>1.0</td>
<td>1.3</td>
<td>2.8</td>
<td>2.8</td>
<td>3.2</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. The response for wt DntR and mutants to a set of potential inducers. The response is measured as fold of induction. Responses that gave at least a 1.5 fold increase in the fluorescence are marked in bold numbers. BEN = benzoic acid, 4-NB = 4-nitrobenzoic acid, 2-NB = 2-nitrobenzoic acid, 2-HT = 2-hydroxytoluene, 2,4-DNT = 2,4-dinitrotoluene, 2-NT = 2-nitrotoluene, 4-NT = 4-nitrotoluene

The data for the inducer responses described in paper I was obtained during non-optimized conditions (optimized in paper II). Therefore, it is possible that some mutations that gave non-significant changes of the inducer specificity profile could have given an observable change in the more sensitive conditions used later. However, reanalysis of the H169V and F111L/H169V mutants gave a similar inducer specificity profile also during these improved conditions (unpublished data).
4.3.2 Improving the response to DNT by directed evolution of DntR

As mentioned in section 4.2.1, optimization of growth conditions and gene arrangement could improve the response to DNT (paper II). However, also with improved conditions, the sensitivity for DNT in wt DntR and “NtdR” is still very low and the response to salicylate is much higher than that to DNT. In the work described in paper III, “NtdR” (analysed in paper II) was selected as a starting point for several rounds of random mutagenesis with selection steps between each round. A pool of mutants with an average improved response to DNT was then subjected to a combination of random mutagenesis and recombination followed by selection. After this step, six clones were selected for recombination into a fifth library (Figure 17).

Figure 17. Directed evolution process and the resulting improvement of the DNT response. Left: Schematic illustration of the selection process in paper III. The numbers stands for each library generation. Library 1-3 were created using epPCR with Mutazyme II polymerase. The library size was in the range of 1-7x10⁷ transformants/library. Library 4 was created using a modified variant of StEP using Mutazyme II polymerase, resulting in recombination of clones selected from library 3 with additional point mutations introduced by Mutazyme II. The best responding clones selected from library 4 were then recombined with StEP, this time using Pfu Polymerase. The success of recombination was confirmed by sequencing of 8 clones from each of library 4 and 5. Right: Overlay histograms from FACS analysis showing the fluorescence for different cell-populations. Top: Cells with wt DntR. Down: Cells with the improved p3c19 variant. Peaks in grey show cells with addition of only solvent (control). Peaks in green show cells with addition of 500 µM DNT. Peaks in red show cells with addition of 500 µM salicylate.
At this stage, the basal level of fluorescence was much higher for a majority of the clones compared to the original templates. By growing the library of clones in modified minimal medium (M9) instead of the rich LB medium, a subpopulation with lowered basal level of fluorescence in M9 could be distinguished. From this subpopulation, several clones with an improved DNT-response could be isolated (Table 2). The best-responding DntR clone “5.3P3c19” displayed a greatly improved response to DNT (Figure 17) and also a broad inducer specificity profile.

<table>
<thead>
<tr>
<th>Protein variant</th>
<th>SAL</th>
<th>DNT</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>wt</td>
<td>10.4</td>
<td>1.2</td>
<td></td>
</tr>
<tr>
<td>5.2R3e18</td>
<td>8.5</td>
<td>4.0</td>
<td>S40A, T46A, N49K, I74V, R165H, P227S, I230N, I232V</td>
</tr>
<tr>
<td>5.3P3c21</td>
<td>12.7</td>
<td>4.8</td>
<td>T46A, N49K, L54V, P114Q, M222I, P227S, I230N, I232V</td>
</tr>
</tbody>
</table>

Table 2. The response to salicylate (SAL) and 2,4-dinitrotoluene (DNT) for clones obtained by directed evolution. Clones that displayed an improved response to DNT (>4-fold) together with the response for wt DntR and NtdR. The response is measured as fold of induction. The substitutions compared to wt DntR are also listed, with substitutions occurring in NtdR marked in bold.

Some substitutions were found to occur in several of the DNT-responding mutants. It is not possible to link these mutations directly to a change of the inducer specificity, but the occurrence of some of them in several of the best-responding clones together with the structural data presented in paper IV makes it possible to attribute a functional role to some of the sites at which mutations were identified.

Some substitutions of residues lining the primary IBC, such as H169D and L151F are expected to affect inducer binding directly. Other enriched substitutions were found in the β9-α9 linker and the α9 helix in RD2 that changes position considerably when the conformation change from open to close as described in paper IV. This region is also at the dimer interface (between two IBDs arranged in a head-to-tail fashion). So, structural changes here are expected to give rise to rearrangements in the whole tetramer.
4.3.3 Model of how a ligand induces DntR-mediated transcriptional activation.

Inducer binding is coupled to a conformational change of the DntR protein into a conformation that is responsible for transcriptional activation. However, not all ligands are necessarily inducers. As seen from the data presented in paper I, fluorescence-quenching measurements showed that salicylate and benzoate have a similar affinity for wt DntR in vitro, yet only salicylate gives a clear transcriptional activation in vivo. Other mutants, however, have the same sensitivity for salicylate as benzoate. These data suggest that a ligand can bind to the primary IBC without acting as an inducer. The hydrophobic inner part of the primary IBC contains many aromatic residues and is large enough to accommodate many different ligands with an aromatic-ring structure. The hydrophilic entrance to the IBC provides possibilities for more specific interactions with functional groups of inducing compounds. Structural data on salicylate binding (paper IV) provide detailed insights on how salicylate interacts with residues lining the primary IBC in DntR. Salicylate forms direct or indirect hydrogen bonds to several residues (such as H169, L151 and T104) that have been substituted in the work presented here.

The position of H169 in the flexible hinge region (Figure 13) together with mutagenesis data (papers I and III) show the importance for this residue in shifting the position of RD1 relative to RD2 into a proper conformation for transcriptional activation. Replacement of this residue with a small hydrophobic residue (L/V) results in mutants with lowered ability to activate transcription [136] (papers I and II). However, the inability for H169L mutants to activate transcription can be restored in combination with other substitutions both at the IBC (paper I) and far away from the primary IBC [136] (paper II). On the other hand, substitution of this residue with a charged aspartate (paper III) or a polar threonine (unpublished data) resulted in mutants with much higher basal level of fluorescence than wt DntR, i.e. transcriptional activation was achieved also in the absence of a ligand.

Several clones (paper III) with an L151F substitution (among many other substitutions) had both broader specificity and higher sensitivity than wt DntR. Also, T104 was substituted in a site-saturation experiment (unpublished results). Interestingly, the substitution T104R on the background of wt DntR yielded a non-inducible mutant with higher basal fluorescence than wt, while T104R on the background of the best-responding mutant “5.3P3c19” (paper III) had low background fluorescence but responded to inducers with an altered inducer response compared to “5.3P3c19”. Also, the substitution T104F on the background of “5.3P3c19” yielded a mutant with high basal fluorescence and higher response to 500 µM DNT than to 500 µM salicylate, with a very broad specificity. This mutant contains two extra aromatic residues in the primary IBC.
(compared to wt DntR) and is induced by aromatic ligands with hydrophilic functional groups at much lower concentrations than wt DntR induction by salicylate (unpublished results).

Figure 18. Model of the DntR-mediated transcriptional activation. Top: DntR is bound to the $P_{DNT}$ promoter also in the absence of inducer. In the absence of inducer, each monomer adopts an open conformation. Four monomers interact to form a tetramer in a compact conformation. In this conformation, the DBD’s interact with the promoter so that the DNA has a high bending angle. Bottom: When an inducer (orange/yellow) binds, RD1(green) and RD2(blue) in each monomer rotate and move towards each other such that each monomer adopts a closed conformation. This changes the interactions at the dimeric interfaces such that the tetramer will adopt an expanded conformation. Probably, IBD interactions in the central part of the tetramers are reduced, while interactions between the linker regions (yellow) play a larger role. In this conformation, the bending angle of the promoter is relaxed, and transcriptional activation occurs.
Altogether, the data suggest that for a ligand to act as an inducer, this ligand must: (i) contain an aromatic ring structure that can interact with aromatic residues in the hydrophobic inner part of the primary IBC, (ii) also contain hydrophilic functional groups that can interact with residues in both RD1 and RD2 and the hinge region (iii) the interaction must also result in moving RD1 and RD2 closer together so that the monomer adopts a closed conformation. Interactions with key residues in both domains mediate the closing and rotation of RD1 relative to RD2. This causes rearrangements of the interface between monomers in the head-to-tail dimer. This is proposed to cause a global conformational change resulting in loss of interactions in the central tetramer region between RD2 regions of opposing monomers so that the tetramer adopts a more expanded conformation. In this expanded conformation, rearrangements of the linker helices are likely to occur such that the position of the DBD pairs change. This conformational change would relax the bending angle of the DntR-bound promoter DNA so that transcriptional activation occurs (Figure 18). However, since there is no structural data of a full-length LTTR in both presence and absence of an inducer, and not in complex with DNA, the detailed mechanism is yet unclear. The picture of how inducer binding leads to transcriptional activation is complicated further by the possibility of a second inducer binding site (paper IV). This putative site displays less specific interactions with salicylate, and could possibly also accommodate other aromatic ligands.

Data from the directed evolution experiments (paper III/unpublished data) also suggest that inducer-independent phenotypes, i.e. DntR variants that are transcriptionally active in the absence of any added inducer, can arise with relative ease without substitutions close to the primary IBC. Also, several broad-specificity mutants that were obtained by directed evolution (paper III) carried no substitutions directly lining the primary IBC (or close to the putative second inducer binding site), but with substitutions in the parts of the IBD that moves considerably when an inducer binds. This illustrates how the induction is determined not only by the direct ligand-protein interaction. The capability for a ligand to act as an inducer is also dependent on the flexibility of the protein, which is important for how readily the protein can adapt the proper conformation for transcriptional activation.

4.4. Concluding remarks and future perspectives

There are several possible improvements to be made for the prototype biosensor. The specificity and sensitivity for detection of DNT needs to be increased. One way to further improve the specificity could be to continue the work with directed evolution of DntR, possibly by site-saturation of key residues. Incorporating the dntR:PDNT:GFP fragment into the chromosome might further improve sensitivity and it would also improve the stability of the reporter system. For
practical applications, incorporating the $\text{dntR:} P_{\text{DNT}}: \text{gfp}$ fragment into a soil bacterium like $\text{Pseudomonas}$ is desirable. Another possibility to improve sensitivity is to develop a cell-free assay.

The work described in this thesis has revealed some clues to mechanisms of inducer binding and transcriptional activation mediated by DntR. However, for a better understanding of the mechanism of transcriptional activation, studies of the interaction of the DntR protein with its promoter are needed. Preliminary DNAse I footprinting (unpublished data) shows protection of the $P_{\text{DNT}}$ promoter at the -35 region (relative to the DntAa-B operon) containing a T-N$_{11}$-A LysR motif (with another LysR motif shifted 2 bp from the first). There is also a T-N$_{11}$-A motif 19 bp upstream of the first T-N$_{11}$-A motif, although no gels with good enough resolution in this region have been obtained. So the details of the binding of DntR to its promoter remain to be investigated.

Also, further investigations regarding the putative second binding site are needed to clarify whether this site truly is biologically significant. Further characterization, especially crystallization, of the DNT-responding mutants obtained by a directed evolution approach could give additional insights about what determines inducer specificity for LTTR proteins. Co-crystallization of the full-length DntR in complex with salicylate would also provide valuable insights into the mechanism of transcriptional activation.
Sammanfattning på svenska

Syftet med denna avhandling har varit att förbättra förmågan hos proteinet DntR att upptäcka DNT. Det långsiktiga målet har varit att använda DntR i en biosensor för att upptäcka sprängämnet TNT, som avger DNT som en "signaturmolekyl". En annan aspekt har varit att bättre förstå den detaljerade mekanismen för hur DntR fungerar.

DntR är ett protein som binder till en viss DNA sekvens (s.k. promotor) och reglerar hur gener intill denna promotor-sekvens läses av (transkriberas). När en inducerande molekyl som t.ex. DNT binder till DntR förändras proteinets struktur på ett sådant sätt att DntR kan aktivera transkription av de gener som finns intill promotor-sekvensen. För att mäta hur DntR reagerar på olika inducerande molekyler har DntR uttryckts i bakterien *Escherichia coli*, som också innehåller promotorn som DntR binder till. Intill promotorn sitter en gen som kodar för proteinet GFP. När en inducerande molekyl binder till DntR, slås transkription av *gfp*-genen på, och det fluorescerande proteinet GFP produceras. Ju mer GFP som produceras i cellerna, desto högre fluorescens kan uppmätas när cellerna analyseras.

I de artiklar som presenteras i avhandlingen har specificiteten för proteinet DntR undersökts, d.v.s. ifall flera olika typer av molekyler (läg specificitet) eller bara en typ av molekyl (hög specificitet) kan binda till DntR. Vidare har vi undersökt hur olika substitutioner i DntR proteinet påverkar specificiteten och sensitiviteten och även hur dessa egenskaper kan påverkas av olika experimentella faktorer.

Effekten av substitutioner har relaterats till strukturdata, där bilder av hur proteinet ser ut på molekylär nivå har tagits fram. Dessutom presenteras även en bild av hur DntR förändras beroende på om inducerande molekyler är bundna eller inte. En sådan strukturbild ökar förståelsen för de mekanismer som gör att bindning av en inducerande molekyl orsakar en förändring av formen hos DntR på så sätt att avläsning av gener kan aktiveras.

Vi har också använt en metod där evolutionära processer härmats för att få fram varianter av DntR med förbättrad respons till DNT. En variant med en drastisk ökning av DNT-responsen har isolerats, och dess egenskaper har karakteriserats.
Acknowledgements

First of all, thanks to Peter Brzezinski, the most encouraging supervisor ever! Always enthusiastic, you have given me the opportunity to try my own ideas.

Irina, your knowledge in biochemical methods and discussions regarding DntR has helped me a lot, thank you!

Our collaborators in France - Cyril and Gordon: thanks for all the help, suggestions and fruitful discussions regarding structure and function.

A big thankyou to my project students; Diana, Theo, Kristin, Leyla, Grigori and Edina. It’s been great fun to supervise and discuss this project with you!

All members, present and past, in the groups of Peter Brzezinski and Pia Ädelroth (Ann-Louise, Christoph, Davinia, Emelie, Gisela, Henrik, Hyun Ju, Håkan, Ida, Kristina, Lina, Linda, Maria, Peter, Ullis and Yafei); thanks for a great time and many nice coffee-break discussions. A special thanks to Joachim and Gustav that had to endure all my mumbling by the computer while writing this thesis.

I would also like to thank everyone else from the department that has contributed to the nice and fun time I’ve had as a PhD student.

Per-Åke Nygren at KTH: thanks for letting me use the FACS vantage, and for discussing all the protein engineering stuff! And thanks to John for the practical help with the FACS.

I would also like to thank SL for providing an excellent working environment. Without the possibility to read and edit on the subway, this thesis would still not have been completed.

Friends from KTH: Diana, Nina, Anna and Emma. Our pizza-nights have been invaluable for discussing everything else about the life as PhD students!

I also want to mention all my other friends that may not have contributed much to this thesis, but nevertheless have contributed to making this time memorable.

Mamma och pappa, tack för att ni alltid stöttat mig.

Till min fina lilla familj: Adde, utan dig skulle avhandlingen aldrig blivit färdig. Jag är glad att du tackade ja till öl den där kvällen för länge sedan, fast du inte ens tyckte om öl 😊 Tuva och Malva, mina fantastiskt underbara ungar, jag kan inte ens föreställa mig hur den här perioden av mitt liv hade sett ut utan er. Jag älskar er!!
References


133. Cormack BP, Valdivia RH, Falkow S: FACS-optimized mutants of the 
134. Hakkila K, Maksimow M, Rosengren A, Karp M, Virta M: Monitoring 
promoter activity in a single bacterial cell by using green and red 
ley AS: Chromosomally located gene fusions constructed in *Acine-
tobacter sp.* ADP1 for the detection of salicylate. *Environ Microbiol*
136. Ju KS, Parales JV, Parales RE: Reconstructing the evolutionary history 
of nitrotoluene detection in the transcriptional regulator NtdR. *Mol 