Design of artificial functional and regulatory systems in bacteria

Von der Fakultät 4: Energie-, Verfahrens- und Biotechnik der
Universität Stuttgart zur Erlangung der Würde eines
Doktors der Naturwissenschaften (Dr. rer. nat.)
genehmigte Abhandlung

Vorgelegt von

Johannes Andreas Heinrich Maier
aus Villingen-Schwenningen, Deutschland

Hauptberichter: Prof. Dr. Albert Jeltsch
Mitberichter: Prof. Dr. Bernhard Hauer
Prüfungsvorsitzender: Prof. Dr. Stephan Nußberger

Tag der mündlichen Prüfung: 25. Juli 2017

Institut für Biochemie und Technische Biochemie der Universität Stuttgart
Abteilung Biochemie

2017

Eidesstattliche Versicherung

Hiermit versichere ich, dass ich diese Arbeit selbst verfasst habe und dabei keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

Stuttgart, 31.03.2017

Johannes Maier

Contents

Acknowledgements

First, I want to thank my supervisor and mentor, Prof. Dr. Albert Jeltsch for giving me the opportunity to work in his lab. He provided great guidance, support and help in all stages of the preparation of this thesis.

I am grateful to Prof. Dr. Bernhard Hauer for being co-referee and Prof. Dr. Stephan Nußberger for being the head of the thesis committee.

I would like to thank Prof. Dr. Christina Wege, Prof. Dr. Monilola Olayioye, and Prof. Dr. Dieter Wolf for taking part in the evaluation procedure.

I want to thank my lunch-break-friends Denis, Peter, Micha, and Phil for interesting conversations about a multitude of topics.

Thanks to all my colleagues in the lab for technical help and a nice working atmosphere.

I also would like to mention my parents and my siblings Jörg and Ute for support during my studies.

In particular, I am very thankful to my wife Lisa for support, patience and love. My gratitude goes also to my lovely daughter Lotte who made my life sleepless and exciting since the day of her birth.

List of Publications

Maier, J.A.H., Möhrle, R., Jeltsch, A. **Design of synthetic epigenetic circuits featuring memory effects and reversible switching based on DNA methylation**. Nature Communications 8: 15336. (2017)

In this study, we developed a bistable synthetic epigenetic memory system that stores the transient signals in DNA methylation patterns. To this end, we constructed a synthetic CcrM-methylation dependent gene expression system in *E. coli*. We set a gene for the methyltransferase CcrM under the regulation of this system and thereby constructed a system with positive feedback. Upon transient trigger signals, the system switches from a stable off-state to a stable on-state. Harnessing the epigenetic nature of the system, we also showed that it is possible to switch the system to the off-state by targeted protein degradation of the CcrM protein.

Maier, J.A.H., Albu, R.F., Jurkowski, T.P. & Jeltsch, A. **Investigation of the C-terminal domain of the bacterial DNA-(adenine N6)-methyltransferase CcrM.** Biochimie 119, 60-67 (2015).

In this paper, we show that the C-terminal domain of *Caulobacter crescentus* CcrM is involved in DNA binding and propose a reclassification of CcrM into the δ-class of DNA-(adenine N6)-methyltransferases.

Maier, J.A.H., Ragozin, S. & Jeltsch, A. **Identification, cloning and heterologous expression of active [NiFe]-hydrogenase 2 from Citrobacter sp. SG in Escherichia coli.** Journal of biotechnology 199, 1-8 (2015).

In this paper, we report the identification of a *Citrobacter* sp. capable of hydrogen production under aerobic cultivation conditions. We identified the active hydrogen producing enzyme to be a type 2 hydrogenase and cloned and expressed this hydrogenase successfully in *E. coli* cells.

Zusammenfassung

In der Synthetischen Biologie wird versucht, lebende Zellen genetisch so zu programmieren, dass sie gewünschte Aufgaben ausführen. Beispiele sind die Produktion von Brenn- oder Treibstoffen, das Töten von Krebszellen oder das Reagieren auf einen bestimmten Umweltstimulus mit einer spezifische biologischen Antwort. In den letzten 30 Jahren sind schnelle und kostengünstige Klonierungsverfahren und Sequenziermethoden entwickelt worden, die grundlegend zum Aufkommen dieses Wissenschaftsfelds beigetragen haben. Auch hatten Wissenschaftler, die aus dem Ingenieurbereich kamen, einen wesentlichen Einfluss auf die Entwicklung des Forschungsgebiets. So wird versucht, mit Komponenten, die in ihrem Verhalten elektrotechnischen Bauteilen ähneln, modulare komplexe Systeme zu bauen. Zur Umsetzung dieser Arbeitsweise ist es wichtig, die einzelnen grundlegenden Komponenten möglichst genau zu verstehen. In dieser Arbeit wurde der epigenetische Mechanismus der DNA Methylierung in einem synthetischen Biologie Ansatz genutzt; epigenetische Modifikationen sind vererbbar, beeinflussen die zu Grunde liegende genetische Information nicht und sind reversibel.

In einem Projekt dieser Arbeit wurde die bakterielle Methyltransferase CcrM aus *Caulobacter crescentus* biochemisch und strukturell untersucht. Bisher gab es nur wenige Strukturinformationen über das Enzym. Interessanterweise zeigen CcrM und homologe DNA-(Adenin N6)-Methyltransferasen die ebenso wie CcrM 5‘-GANTC-3‘ Zielsequenzen methylieren, eine konservierte C-terminale Domäne auf, dessen Funktion jedoch nicht bekannt war. Hier wurde gezeigt, dass die C-terminale Domäne von CcrM essenziell für die DNA Bindung ist. Dieses Ergebnis deutet darauf hin, dass CcrM eine Proteinarchitektur aufweist, die theoretisch schon beschriebene wurde, jedoch wurde noch keine solche Methyltransferase gefunden. Dies erfordert eine neue Einordung von CcrM in der Klassifizierung von DNA-(Adenin N6)-Methyltransferasen. CcrM stellt das erste beschriebene Beispiel der δ-Klasse dar.

In einem weiteren Projekt wurde die CcrM Methyltransferase als Schlüsselenzym zur Entwicklung eines synthetisch-epigenetischen Memorysystems verwendet. Zusammen mit einem designten Zink Finger Protein, das als methylierungssensitiver transkriptioneller Repressor fungiert, wurde ein bistabiles System mit positivem Feedback entwickelt. CcrM reguliert hierbei die Transkription seines eigenen Gens

durch Methylierung der Promoterregion. Sobald CcrM einmal exprimiert wurde, bleibt das System durch die kontinuierliche Methylierung der Promoterregion dauerhaft im angeschalteten Zustand. Das System kann durch eine transiente Temperaturerhöhung, die transiente Anwesenheit des Zuckers Arabinose oder transiente DNA-schädigende Bedingungen vom ausgeschalteten Zustand in den angeschalteten Zustand gebracht werden. Die vorübergehenden Signale werden dann dauerhaft in sich teilenden Bakterien in Form von DNA Methylierungsmustern gespeichert. Das entwickelte System könnte als Biosensor System, Biocontainment System oder zum Beispiel als industrielle Proteinexpressionsplattform Anwendung finden. Die Entwicklung ähnlicher Systeme, die andere Methyltransferasen mit anderen Methylierungsmustern zur Informationsspeicherung nutzen, würde es erlauben mehrere Inputsignale gleichzeitig zu speichern und Informationen mit Boolescher Logik zu verarbeiten. Dies würde auch die Entwicklung von komplexen Biocomputing Systemen ermöglichen.

In einem weiteren Projekt, wurde ein Bakterium der *Citrobacter* Gattung entdeckt, das in der Lage ist, unter aeroben Kultivierungsbedingungen Wasserstoff zu produzieren. Das verantwortliche Enzym wurde als eine Typ 2 Hydrogenase identifiziert und aktiv in *E. coli* exprimiert. Dieses ungewöhnliche Enzym könnte ein Baustein sein, um Biowasserstoff gekoppelt an Photosynthese herzustellen.

Abstract

Synthetic biologists aim to program living cells to produce fuels, chemicals or pharmaceutical products, accomplish specific tasks, like attacking cancer cells, or reacting to certain environmental stimuli. The advent of easy and fast cloning techniques and DNA synthesis as well as steadily faster and cheaper DNA sequencing allowed for fundamental developments in the field of synthetic biology over the last 30 years. Engineers have strongly influenced the field of synthetic biology by means of applying engineering principles to modify and equip cells to perform desired functions. In order to be able to work like an engineer by assembling certain genetic parts and thereby creating new biological functions, it is of great importance to have profound knowledge of the used biological parts. Here, the epigenetic mechanism of DNA methylation was combined with synthetic biology approaches; epigenetic modifications are heritable but reversible and do not alter the underlying DNA sequence.

In this thesis, first, the bacterial DNA methyltransferase CcrM was characterized structurally and biochemically. Due to a lack of structural data, little is known about DNA binding and sequence recognition of CcrM. CcrM and homologous DNA-(adenine N6)-methyltransferases that methylate adenines at 5'-GANTC-3' sites, exhibit a conserved C-terminal domain with unknown function. Here, it was shown that the C-terminal domain of CcrM is involved in DNA binding. This result suggests that CcrM exhibits a protein architecture that has been theoretically described but has not been discovered in a bacterial methyltransferase. This demands for a reclassification of CcrM to the δ-class of bacterial DNA-(adenine N6)-methyltransferases.

Next, the methyltransferase CcrM was used as a corner stone together with a designed DNA methylation sensitive zinc finger protein in the development of a synthetic epigenetic memory device. The zinc finger protein acts as a transcriptional repressor and its binding can be modulated by CcrM introduced DNA methylation. Using these two building blocks, a bistable system was created in which the zinc finger repressor regulates a gene for CcrM. Once CcrM is expressed, the system is durably switched on due to the positive feedback. The iteratively developed system was designed to react to transient sensory information like heat, presence of arabinose, or DNA damaging conditions and stores this information durably but also reversibly in DNA methylation patterns in living bacteria. This system can find use in biotechnology

applications e.g. as biosensors, biocontainment systems, or as an industrial protein induction platform. The development of similar systems using complementary methyltransferases will allow for multiplexing and processing of multiple input signals, enabling Boolean logic operations and more complex biocomputing applications.

Additionally, a *Citrobacter* species expressing an unusual bacterial type 2 hydrogenase was discovered that is capable of hydrogen production under aerobic cultivation conditions. The unusual enzyme was actively expressed in *E. coli* and might be a valuable biological part for one-step biohydrogen production coupled to photosynthesis.

List of abbreviations

4mC	N4-methyl-cytosine
5mC	C5-methyl-cytosine
6mA	N6-methyladenine
AdoMet	S-adenosyl-L-methionine
AI	after induction
BI	before induction
Cas9	CRISPR associated protein 9
CcrM	cell-cycle regulated DNA methyltransferase
Cori	chromosomal replication origin in Caulobacter
CRISPR	clustered regularly interspaced short palindromic repeats
CRISPRa	CRISPR mediated gene activation
CRISPRi	CRISPR interference
CTD	C-terminal domain
CtrA	cell cycle transcriptional regulator A
Dam	deoxyadenosine DNA methyltransferase
dCas9	catalytically inactive Cas9 protein
DIC	differential interference contrast
EGFP	enhanced green fluorescent protein
EMSA	electro mobility shift assay
GFP	green fluorescent protein
MAGE	multiplex automated genome engineering
MTase	methyltransferase
MV	methyl viologen
NTD	N-terminal domain
ppm	parts per million
PTM	post-translational modification
RM	restriction/modification
SDS PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SMRT	single molecule real time sequencing
TALE	transcription activator-like effectors
TBT	TATA box-binding protein
TRD	target recognition domain
ZnF	zinc finger
ZnF4	tetrameric zinc finger repressor

1 Introduction

This work deals with the development of a synthetic epigenetic memory system in *Escherichia coli* that can sense and memorize different transient input signals in form of DNA methylation patterns, the biochemical and structural investigation of the bacterial DNA-(adenine N6)-methyltransferase CcrM from *Caulobacter crescentus*, and the identification and heterologous expression of the oxygen tolerant type 2 hydrogenase from *Citrobacter freundii*. Relevant topics will be introduced in the next chapters of this section.

1.1 Synthetic biology and synthetic circuit design

1.1.1 An overview of synthetic biology

In the year 1961, upon their investigations on the lac operon, François Jacob and Jacques Monod postulated the existence of regulatory circuits that can respond to environmental inputs (Monod and Jacob 1961). They hypothesized the assembly of new circuits (Jacob and Monod 1961) and thereby originated the field of synthetic biology.

Today, synthetic biologists desire to rationally program living cells in order to understand functional parts, control cellular behavior, and create new biological systems and functions, which later could be of technical or industrial use. They apply engineering principles to predictively construct new functions by the assembly of biological parts in novel ways. Hereby, they use either naturally occurring parts or newly engineered bioparts and molecular biology tools.

The development of synthetic biology is based on the emergence and improvements of molecular biology techniques during the 70s, 80s, and 90s of the last century. These include molecular cloning techniques, PCR, and automated DNA sequencing. In the year 2000, two groups reported the construction and engineering of two synthetic genetic circuits, a “toggle switch” and a “repressilator”, that are now often considered as the cornerstones of the field of synthetic biology (Elowitz and Leibler 2000; Gardner

et al. 2000) (Figure 1). The toggle switch is a gene circuit that exhibits a bistable behavior and consists of two repressor genes that negatively regulate each other. Switching between the two stable states is achieved by transient signals, like chemical compounds or thermal induction. The repressilator is a synthetic gene network consisting of three transcriptional repressor genes. Each expressed repressor protein represses the transcription of the consecutive repressor gene resulting in an oscillating expression profile of a reporter gene.

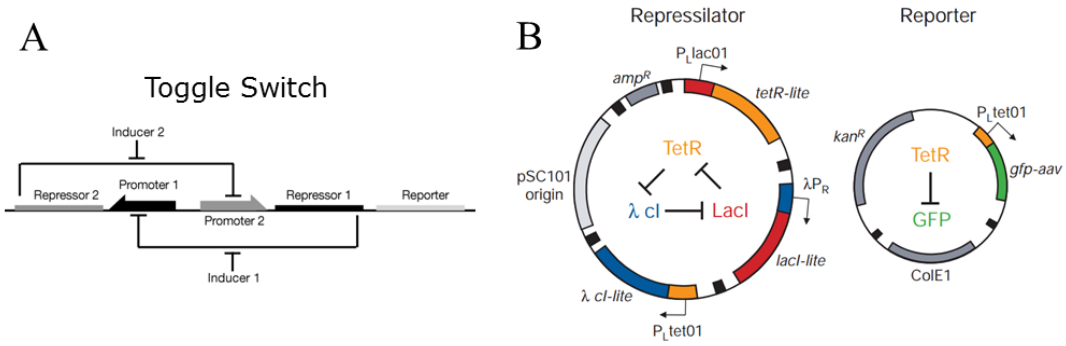


Figure 1. Seminal synthetic biology gene circuit designs. A) The toggle switch that can stably switch between to states. (Picture taken from (Gardner et al. 2000).) B) The repressilator exhibits an oscillating protein expression profile of three repressor proteins. (Picture taken from (Elowitz and Leibler 2000).)

In the following years, the engineered systems became more and more diverse and versatile and an increasing number of control elements have been used. Additional to regulation at the transcriptional level, RNA-based systems that are regulated at the level of translation have been developed (Isaacs et al. 2004). Also prokaryotic cell to cell communication systems (quorum sensing) have been incorporated into the construction of synthetic circuits resulting in intercellular gene circuits. A primary example circuit resulted in a 2D pattern formation in *E. coli* (Basu et al. 2005). Later, quorum sensing based circuits have been used to manipulate bacteria to lyse synchronously in pulsatile patterns, which was coupled to the release of a therapeutic payload and may find application in live drug delivery systems (Danino et al. 2010; Din et al. 2016). Among many other synthetic circuits, biological counting devices have been devolved that are based on riboregulators or recombinases (Friedland et al. 2009).

The introduction of new cloning techniques like Gibson Assembly has leveraged the constraints of restriction endonuclease based cloning methods and, together with

better and cheaper gene synthesis, it has become easier to build and assemble gene circuits from scratch (Gibson et al. 2009). DNA synthesis and assembly methods have facilitated the synthesis of whole bacterial genomes, a mouse mitochondrial genome and a yeast chromosome (Gibson et al. 2010a; Gibson et al. 2010b; Annaluru et al. 2014). Furthermore, with the method of multiplex automated genome engineering (MAGE) it is now possible to simultaneously modify prokaryotic genomes at multiple sites with high efficiency (Wang et al. 2009a). Researchers at the J. Craig Venter institute built synthetic minimal genomes (smallest 531 kb) based on the reduction of the *Mycoplasma mycoides* genome, exemplifying the creation of semisynthetic life forms (Gibson et al. 2010a; Hutchison et al. 2016).

The field of synthetic biology is not restricted to bacteria. Synthetic biology approaches have also found their way into eukaryotic systems, like mammalian model cell systems and plants (Gersbach et al. 2016; Nemhauser and Torii 2016). Especially the emergence of programmable DNA binding devices like zinc finger proteins, TALEs (transcription activator-like effectors), and the CRISPR Cas9 system (clustered regularly interspaced short palindromic repeats, CRISPR associated protein 9) have led to a massive progress in the field (Weber and Fussenegger 2012; Lienert et al. 2014).

The mentioned examples of synthetic biology devices are mainly proof-of-principle studies, however, especially in the field of metabolic engineering, synthetic biology along with systems biology intensified progress and now show real industrial applications in the production of bulk chemicals, fine chemicals, drugs and fuels (Keasling 2010; Lee et al. 2012). The production of the antimalarial drug artesimin is a paradigm for the industrial application of synthetic biology approaches (Ro et al. 2006).

1.1.2 Building blocks for genetic circuit construction

Using cells as biosensors, input signals can, for example, be inorganic or organic chemicals, osmolarity, physical inputs like temperature, ionizing radiation, or light. Any desired output signals can be coupled to the input via genetic circuits. Examples are the expression of reporters like GFP (green fluorescent protein), the synthesis or secretion of a therapeutic compound, triggering of a particular cellular behavior like

biofilm formation, or programmed cell death. Input signals can be processed at the level of transcription, translation, and via post-translational systems.

The basic level of circuit regulation and biosensing is transcriptional regulation that functions by modifying RNA polymerase access to target promoter sequences or by modifying progression of RNA polymerase. Biological parts that influence RNA polymerases are, for example, DNA binding proteins (see chapter 1.1.3) or recombinases that invert, excite, or insert DNA sequences (Roquet et al. 2016). Most of the classical inducible protein expression systems are examples of natural transcriptional biosensors that modulate transcription via binding in the promoter region of their target genes (Lutz and Bujard 1997). Additionally, synthetic variations of theses transcription factors have been developed and used in synthetic biology approaches, in which DNA binding domains and regulatory domains have been reassembled in novel ways (Meinhardt et al. 2012; Chan et al. 2016).

Riboswitches can regulate gene expression at transcriptional and translational level (Serganov and Nudler 2013). In case of transcriptional regulation, ligand binding dependent mRNA conformations can act as terminators or antiterminators by the formation of hairpin structures or by preventing them. In case of riboswitches that act on the translational level, the regulated mRNA contains a 5’ untranslated region that comprises an aptamer sequence that can bind small molecules and change its folding (or three-dimensional structure) upon binding. The structural change can then regulate ribosomal translation (Waters and Storz 2009). Synthetic design of aptamers allows customization of riboswitches (Berens et al. 2015). While riboswitches are regulated in cis by RNA-RNA interactions, there are also small RNAs that can act in trans in the process of translational regulation (Qi and Arkin 2014). In natural systems, small RNAs are involved in gene regulation under low iron conditions or oxidative stress, for example (Waters and Storz 2009). Synthetic systems that sense RNAs have also been developed (Green et al. 2014).

Post-translational regulation uses protein receptors to sense signals and start a signal cascade that provokes a cellular response (Khalil and Collins 2010). Either this can result in solely post-translational circuits (Olson and Tabor 2012 229) or the signal can provoke response at transcriptional or translational level (Skerker et al. 2008 233).

1.1.3 DNA binding proteins

DNA binding proteins are one key component of virtually all artificial genetic circuits. They can be adapted from natural systems, modified, or repurposed in new settings. Examples include, the TetR and homologs (Stanton et al. 2014), the λ phage repressors Cro or CI (Oppenheim et al. 2005), the LacI repressor and related repressors (Zhan et al. 2010), as well as synthetic chimeric transcription factors of this family in which DNA binding domains and regulatory domains were interchanged (Meinhardt and Swint-Kruse 2008; Meinhardt et al. 2012). More recently, catalytically inactive Cas9 protein (dCas9) has been shown to be applicable as prokaryotic transcriptional repressor by blocking initiation or elongation of RNA synthesis also referred to as CRISPR interference (CRSPRi)(Qi et al. 2013). Additionally, transcriptional activation using dCas9 fusion proteins was revealed to be possible in bacteria and is called CRISPR mediated gene activation (CRISPRa)(Bikard et al. 2013). In eukaryotic systems, both, gene activation and repression has also been shown to be conceivable by different means (Konermann et al. 2014; Dominguez et al. 2015). Of note, there is a major difference between DNA interaction of conventional transcription factors and systems based on CRISPR Cas9 system. Conventional bacterial and eukaryotic transcription factors interact with DNA and recognize DNA sequence mainly via protein-DNA contacts and bind their target site via DNA binding domains like zinc finger domains, helix-turn-helix motifs, leucine zippers, or others (Rohs et al. 2010). In contrast, the Cas9 protein binds DNA via a guide RNA and the sequence recognition and binding is based on Watson/Crick base pairing between the guide RNA and one DNA strand, resulting in an Cas9-gRNA-DNA complex (Nishimasu et al. 2014). Hence, DNA modifications like DNA methylation do not play a major role in modulation the binding behavior. However, for example zinc finger proteins or TALEs form DNA contacts via the major groove of DNA and hence DNA methylation can modulate DNA binding (Sasai et al. 2010; Deng et al. 2012; Liu et al. 2012; Liu et al. 2014)(Figure 2).

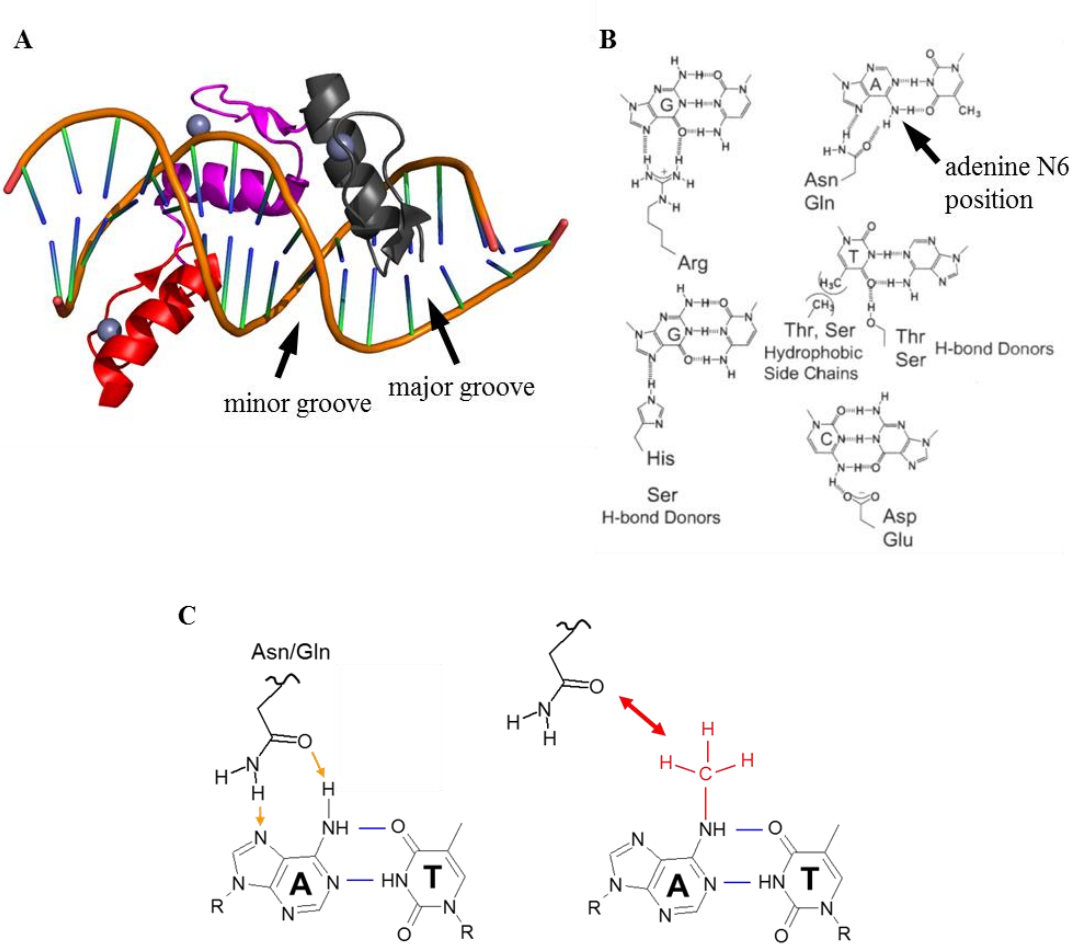


Figure 2. DNA binding and sequence recognition by zinc finger proteins. A) Structure of zinc finger protein in complex with DNA (PDB ID: 2KMK, (Lee et al. 2010)). In red, cyan, and grey individual zinc finger domains are indicated with a complexed zinc ion (light grey sphere). Zinc finger proteins make DNA contacts in the major groove of DNA. B) Interactions of amino acid side chains of zinc finger proteins with DNA bases via H-bonds and hydrophobic interactions. In theory, methylation at the N6 position of adenine bases interferes with amino acid base interactions. In contrast, there is also a report on a plant GATA zinc finger protein that shows enhanced binding to DNA that carries adenine N6 methylation (Sugimoto et al. 2003). (Picture taken from (Sera and Uranga 2002).) C) Interaction of asparagine or glutamine residues in zinc finger proteins with AT base pairs. This interaction is blocked by adenine-N6 methylation (Aggarwal et al. 1988; Wolfe et al. 2000). (Panel C is taken from Supplementary Fig. 2A from Appendix 2.)

1.1.4 Limitations and adjustments of synthetic circuits

Synthetic circuit design can be used to program cells in a rational way. By combination of several simple circuits, more complex systems can be implemented, which integrate different inputs and apply a kind of Boolean logic (Tamsir et al. 2011; Wang et al. 2011). However, there are certain limitations of combining different gene circuits that range

from insufficient dynamic ranges, context dependence of certain genetic parts, transcriptional read-through to crosstalk of regulators. Additionally, potential interactions with host cells have to be considered (Brophy and Voigt 2014). Adjustments to circuits can be made e.g. by copy number adaptations of the circuit's plasmids or the genomic integration of the circuit, promoter strength variation, ribosomal binding site strength alteration, mRNA lifetime variation, or targeted protein degradation.

During this work, a zinc finger protein has been engineered to bind DNA in an adenine-N6 methylation dependent manner and was used as an artificial transcriptional repressors in *Escherichia coli*. Applying this repressor in synthetic circuit design together with a bacterial methyltransferase, that is able to modulate DNA binding of the repressor, a synthetic epigenetic memory system has been established in *Escherichia coli*.

1.2 DNA methylation and Epigenetics

DNA methylation is found in both prokaryotes and eukaryotes and adds additional information to DNA without altering the DNA sequence (Cheng 1995). In bacteria, three types of DNA methylation occur, that do not interfere with base pairing. These are 5-methylcytosine (5mC), N4-methylcytosine (4mC) and N6-methyladenine (m6A) (Jeltsch 2002) (Figure 3). In each case, the methyl group protrudes into the major groove of the DNA helix and it can be read by DNA binding proteins (see also 1.1.3). In mammals, DNA methylation mainly referrers to methylation of cytosines at the C5 position in 5'-CG-3' sequences. However, recently, N6-methyladenine was also discovered as a mammalian DNA modification, but its biological role is still unclear (Luo et al. 2015; Wu et al. 2016). Cytosine methylation at the C5 position plays crucial roles in cell development and differentiation and often is involved in the onset and progression of diseases (Bergman and Cedar 2013). In particular, promoter methylation leads to gene repression, whereas methylation in gene bodies is observed in highly transcribed genes (Baylin and Jones 2011; Jones 2012). Furthermore, DNA methylation is involved in genomic imprinting (Smith and Meissner 2013; Horsthemke 2014), X-chromosome inactivation (Gendrel et al. 2012) and the silencing of transposable elements (Smith and Meissner 2013). DNA methylation is one of the best

studied epigenetic marks in mammals and is introduced by the two *de novo* methyltransferases Dnmt3a and Dnmt3B and maintained by Dnmt1 (Okano et al. 1999; Jeltsch 2006). Histone post-translational modifications (PTMs) and non-coding RNAs function in concert with DNA methylation and encode the epigenome of the cell. Epigenetic processes convey inherited information for gene expression profiles by regulation of chromatin organization (Jaenisch and Bird 2003; Allis and Jenuwein 2016; Henikoff and Greally 2016). In bacteria, there are certain phenomena that resemble epigenetic mechanisms in higher organisms like heritable DNA methylation patterns involved in regulation of the *pap* operon and the *agn43* gene, as well as IS10 and traJ regulation (Wion and Casadesús 2006).

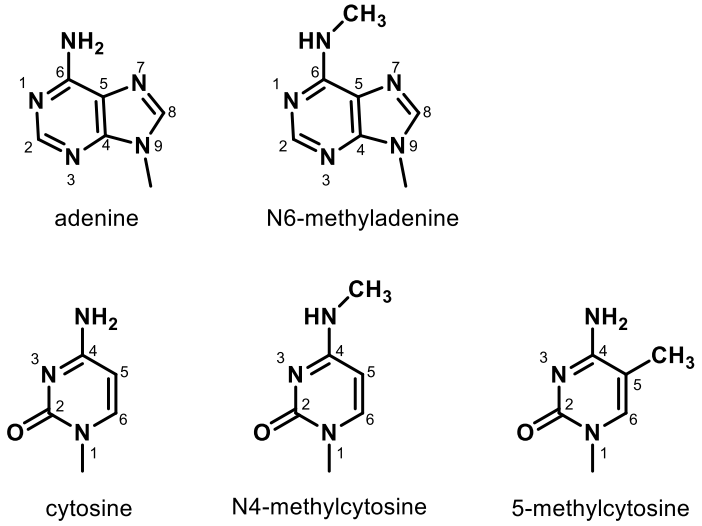


Figure 3. DNA bases with its biological relevant methylated variants.

1.2.1 DNA methylation in prokaryotes

In prokaryotes, most of the known DNA methyltransferases (MTases) are part of restriction/modification (RM) systems. These systems serve to protect prokaryotes from bacteriophage infections. RM systems employ a restriction endonuclease and a DNA MTase with both enzymes recognizing the same target DNA sequence. Usually these DNA target sites are 4 to 8 base pair palindromic sequences. The restriction endonuclease cleaves the DNA at the target sequence only in an unmethylated state, as found on incoming phage DNA during the early steps of infection. The host cell DNA, however, is kept in a methylated state by the corresponding DNA MTase and hence is protected from cleavage (Arber and Dussoix 1962; Bickle and Krüger 1993;

Pingoud and Jeltsch 2001). These systems are complemented by the adaptable CRISPR-Cas systems that have been discovered only recently as an additional defense system of bacteria and archaea against invading phages, independent of DNA methylation (Bhaya et al. 2011).

Aside from MTases of RM systems, there are so called solitary or orphaned MTases that do not serve in protection of DNA from cognate endonucleases. The *Escherichia coli* deoxyadenosine DNA methyltransferase (Dam) and the *Caulobacter crescentus* cell-cycle-regulated methyltransferase (CcrM) are two well characterized examples of this kind (Jeltsch 2002). *E. coli* Dam methylates adenine residues in 5'-G**A**TC-3' sequences and is involved in DNA mismatch repair, initiation of chromosome replication and regulation of gene expression, including the mentioned pap phase variation in uropathogenic *E. coli* (Marinus and Morris 1973; van der Woude et al. 1996; Low et al. 2001; Løbner-Olesen et al. 2005; Marinus and Casadesus 2009). *Caulobacter crescentus* CcrM methylates adenine residues in 5'-G**A**NTC-3' sequences and it is essential for *C. crescentus*, at least under certain growth conditions and plays a central role in regulation of the cell cycle (see chapter 1.2.6) (Reisenauer et al. 1999a; Marczynski and Shapiro 2002; Wion and Casadesús 2006; Gonzalez and Collier 2013).

1.2.2 Classification of bacterial DNA MTases

Bacterial DNA MTases can be classified by the position at which they methylate nucleobases. One class methylates pyrimidine ring carbons, resulting in the formation of a C-C bond in C5-methylcytosine. The other class methylates exocyclic amino nitrogen atoms. They can act on adenine or cytosine bases, resulting either in N6-methyladenine or N4-methylcytosine (Figure 3) (Malone et al. 1995; Jeltsch 2002).

Structurally, both classes of bacterial DNA MTases are similar. They contain a catalytic domain comprising the active site and an AdoMet-binding region, and they contain a target recognition domain (TRD) (Malone et al. 1995; Jeltsch 2002).

Bacterial cytosine-C5 MTases like M.HhaI and M.HaeIII contain ten conserved motifs in the same order in the catalytic domain, N-MTases like M.PvuII or M.TaqI reveal different orders of conserved motifs. By means of the position of the TRD and the order of the conserved motifs in the catalytic domain, N-MTases were classified into classes

called α, β, and γ (Figure 4) (Malone et al. 1995). In this nomenclature, the cytosine-C5 MTases all fall into group γ. Theoretically, there are also the classes δ, ε, and ζ, however, there were no examples found for these classes so far. The different motif arrangements are thought to be the result of circular permutations of the genes of the methyltransferases (Jeltsch 1999). Of note, motifs I, II, III, and X built up the AdoMet-binding region, motifs IV, V, VI, VII, and VIII comprise the active site subdomain (Cheng et al. 1993; Labahn et al. 1994; Schluckebier et al. 1995).

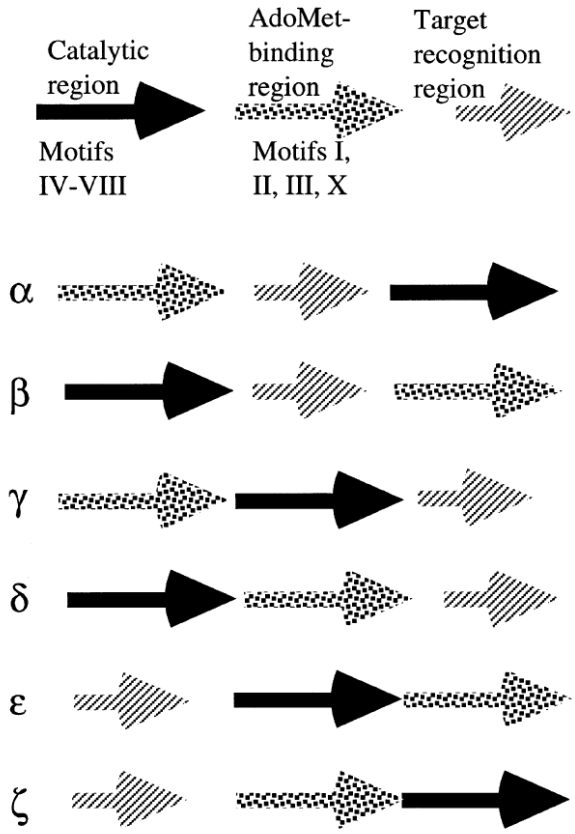


Figure 4. Different arrangements of catalytic domain, AdoMet binding region and TRD in bacterial DNA MTases. Arrangements δ, ε, and ζ are hypothetical. (The image is taken from (Malone et al. 1995).)

1.2.3 Mechanisms of DNA recognition

Binding of specific DNA binding proteins to their target sequences leads to the energetically best possible number of contacts between protein and DNA backbone and specific base pairs. One can discriminate two mechanisms, both leading to the recognition of the specific target sequence (Garvie and Wolberger 2001). Direct read out refers to specific protein DNA target sequence interactions and indirect read out

refers to the recognition of sequence dependent changes in the shape of DNA (Rohs et al. 2010). Structural features of the DNA evoked by the respective DNA target sequence are read by DNA binding proteins via the phosphate backbone, independent of direct interactions of amino acids with the actual nucleotide sequence (Travers 1989; Luscombe et al. 2001). Once the target sequence has been found, specific contacts can be formed additionally and the ΔG release can be used for example to aid conformational changes in the protein, necessary for catalysis. Indirect read-out has been shown to contribute to DNA target site recognition for example for the Trp repressor, the EcoRV endonuclease, or M.EcoRV (Otwinowski et al. 1988; Taylor et al. 1991; Jurkowski et al. 2007).

1.2.4 Catalytic mechanism of exocyclic DNA Methyltransferases

DNA MTases transfer methylgroups from S-adenosyl-L-methionine (AdoMet) (Figure 5) onto their target base. In order to find their target sequence, DNA Mtases bind unspecifically to DNA and diffuse linearly along the DNA until the target sequence has been found. For accessing the target base and to perform the methylation reaction, the base is flipped out of the DNA helix and as a result intrudes into the catalytic pocket of the DNA MTase. Base flipping occurs at the sugar-phosphate backbone and requires breaking of the Watson/Crick hydrogen bonds, but no breaking of covalent bonds (Jeltsch 2002). Methylation reactions are alkylation reactions and require AdoMet as a donor of the methylgroup (Cheng 1995; Jeltsch 2002). AdoMet features a methylthiol moiety that reacts easily with polarizable nucleophiles due to its charged sulfur atom (Cheng and Roberts 2001). The catalytic mechanism of the AdoMet dependent adenine-N6 MTase M.TaqI is shown exemplarily in Figure 6 (Goedecke et al. 2001; Jeltsch 2002). M.TaqI methylates the exocyclic N6 of adenine in 5'-TCG**A**-3' sequences. Once the adenine is flipped out, the target nitrogen gets polarized and becomes a nucleophile by formation of hydrogen bonds to aspartate and proline in the active site. Aspartate and proline are part of the conserved (D/N/S)PP(Y/F) motif present in the active site of adenine-N6 and cytosine-N4 Mtases. Then, the nucleophile attacks the methylthiol resulting in N6 methylated adenine and S-adenosyl-L-homocysteine.

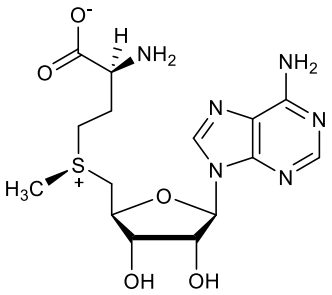


Figure 5. S-adenosly-L-methionine (AdoMet). The methylgroup bound to the charged sulfur can be transferred by MTases resulting in S-adenosyl-L-homocysteine and a methylated DNA base.

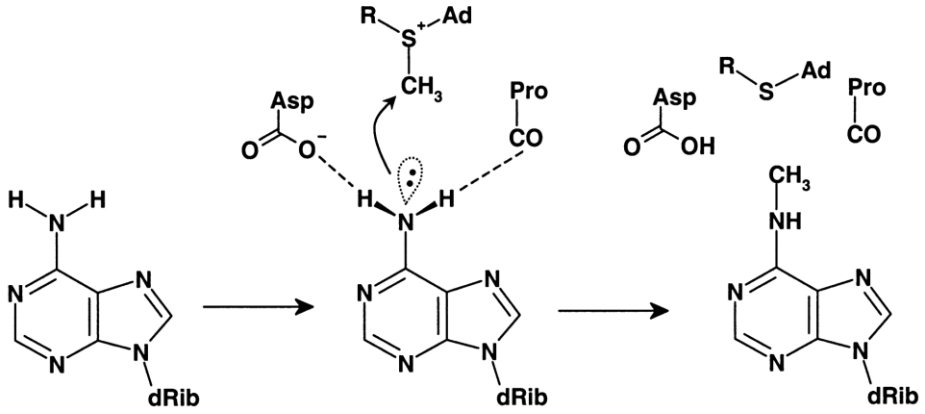


Figure 6. Catalytic mechanism derived from M.TaqI crystal structure. Aspartate and proline sidechains form hydrogen bonds with the nitrogen at the exocyclic N6 position. This leads to a transition from sp^2 to sp^3 hybridization of the nitrogen resulting in a nucleophile, which can attack the methyl group of SAM in a SN2 reaction. (The image is taken from (Jeltsch 2002).)

1.2.5 *Caulobacter crescentus* cell cycle regulated methyltransferase (CcrM)

The cell cycle regulated methyltransferase (CcrM) from *Caulobacter crescentus* is a solitary MTase and plays a crucial role in the *C. crescentus* cell cycle (see Chapter 1.2.6). The adenine-N6 methyltransferase methylates 5'-G**A**NTC-3' sequences in a distributive manner, meaning that CcrM dissociates from the DNA after one methylation reaction (Albu et al. 2012a). In contrast, processive enzymes stay bound to the DNA after a methyl group transfer and slide along the DNA until they find another target site. Moreover, CcrM shows a slight for hemimethylated substrates compared to unmethylated substrates (Albu et al. 2012a). There is evidence that CcrM can form dimers (Shier et al. 2001), however, due to the lack of a structural data, it is unclear how this dimerization is accomplished and where intermolecular interactions might

appear. Until recently, CcrM had been classified as a β-class MTase (Gonzalez et al. 2014). CcrM reveals an additional C-terminal domain with unknown function that is not part of the consensus MTase fold, but conserved among CcrM homologs.

1.2.6 The cell cycle of *Caulobacter crescentus*

Caulobacter crescentus is an aquatic Gram-negative alpha-proteobacterium that divides into two morphologically distinct daughter cells, one stalked sessile cell and one mobile swarmer cell (Curtis and Brun 2010). Only the stalked cell can initiate DNA replication followed by cell division. The swarmer cell first has to attach to a substrate and differentiate into a stalked cell in order to start replication of the chromosome and to divide (Marczynski and Shapiro 2002; Collier 2012). In contrast to *Escherichia coli*, *Caulobacter crescentus* replicates its chromosome exactly once per cell cycle (Roberts and Shapiro 1997; Marczynski 1999). The cell cycle of *Caulobacter crescentus* is tightly controlled by three transcription factors (DnaA, GcrA and CtrA) and the cell cycle regulated methyltransferase CcrM (Curtis and Brun 2010).

In the stalked cell, DNA replication starts with binding of DnaA to the *Caulobacter* chromosomal origin (Cori), which is fully methylated at 5'-GANTC-3' sites that are located there. DnaA starts to unwind DNA in an ATP dependent manner (Skarstad and Boye 1994). Additionally, DnaA acts as a transcription factor and stimulates the transcription of *gcrA*. GcrA in turn regulates many cell cycle regulated genes, such as genes involved in DNA replication (Collier et al. 2006; Collier 2012).

During DNA replication the *Caulobacter* chromosome becomes hemimethylated and the duration of this state is dependent on the individual location of the respective 5'-GANTC-3' site (Zweiger et al. 1994; Kozdon et al. 2013). The *dnaA* gene's promoter region becomes hemimethylated immediately after replication initiation, as it is located close to the Cori on the chromosome. The activity of the *dnaA* promoter is epigenetically regulated by its methylation state; it is active in the methylated state and inactive in hemimethylated state. Hence, *dnaA* transcription diminishes, once the DNA replication fork has passed the *dnaA* promoter (Collier et al. 2007). However, it should be noted that there are conflicting reports claiming, that *dnaA* regulation could also occur independent of the methylation state of 5'-GANTC-3' sites in the promoter of *dnaA* (Cheng and Keiler 2009; Jonas et al. 2011; Gonzalez et al. 2014). Afterwards,

the protein HdaA inactivates existing DnaA and DnaA is degraded by the protease ClpP (Gorbatyuk and Marczynski 2005; Collier and Shapiro 2009). This contributes to the prevention of multiple initiations of chromosomal replication.

Furthermore, DNA replication is negatively regulated by the phosphorylated cell cycle transcriptional regulator A (CtrA) through competitive binding to the Cori (Collier 2012). Once the promoter of the *ctrA* gene has been replicated and converted into the hemimethylated state, *ctrA* transcription is activated by GcrA binding to one of the two promoters of the *ctrA* gene (Reisenauer and Shapiro 2002). Subsequently, expressed CtrA activates its own transcription via the strong second promoter, resulting in high CtrA levels (Domian et al. 1999). CtrA in turn represses *gcrA* transcription (Holtzendorff et al. 2004). CtrA in its activated form is present in swarmer and pre-divisional cells and is involved in the regulation of many additionally cell cycle regulated genes, such as *ccrM* (Laub et al. 2002). *ccrM* transcription is activated on the one hand by CtrA, on the other hand by hemimethylation of two 5'-GANTC-3' sites in its promoter (Reisenauer et al. 1999b; Collier et al. 2007). Via this regulation, CcrM is only expressed and active at the end of S-phase prior to cell division and and then reestablishes full methylation at 5'-GANTC-3' sites (Zweiger et al. 1994). Methylation of its own promoter and degradation of CcrM by the Lon protease restricts the activity to this narrow time window (Wright et al. 1996).

During this work, the DNA-(adenine N6)-methyltransferase CcrM from *Caulobacter crescentus* was investigated biochemically and structurally. It was found that the C-terminal tail of CcrM plays a role in DNA binding and thus a re-classification of CcrM to the δ-type DNA-(adenine N6)-methyltransferase has been proposed.

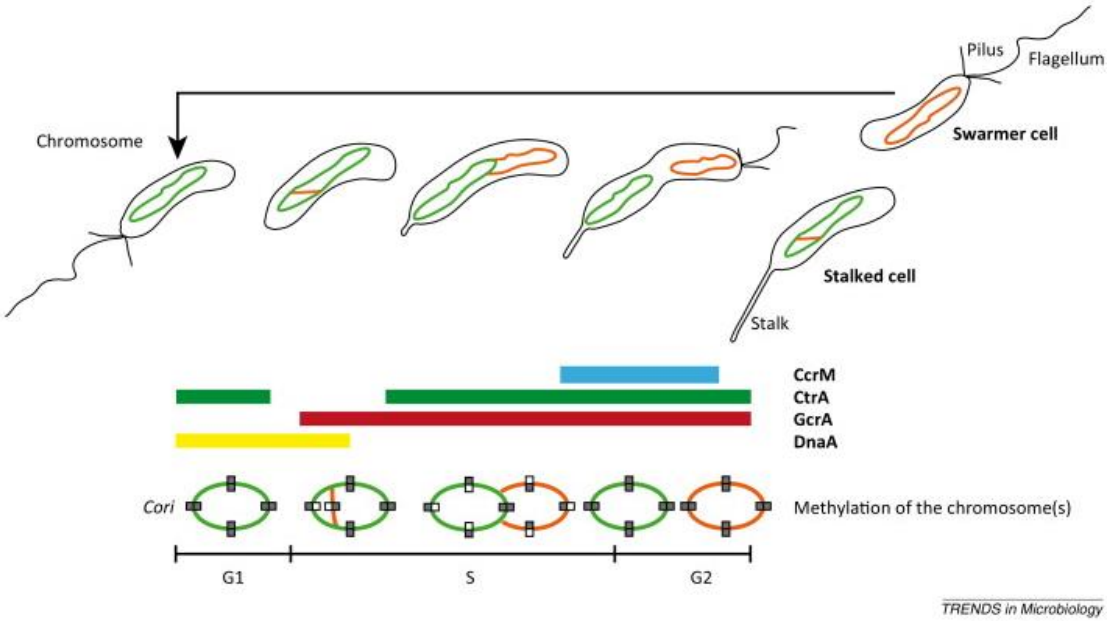


Figure 7. Schematic cell cycle of *Caulobacter crescentus* resulting in a swarmer and a stalked cell. CcrM, CtrA, GcrA and DnaA expression in different phases of the cell cycle is depicted in colored bars. Methylation of the chromosome is illustrated by filled double squares, hemimethylated sites after DNA replication are shown in one filled and one empty square. (Picture taken from (Mohapatra et al. 2014).)

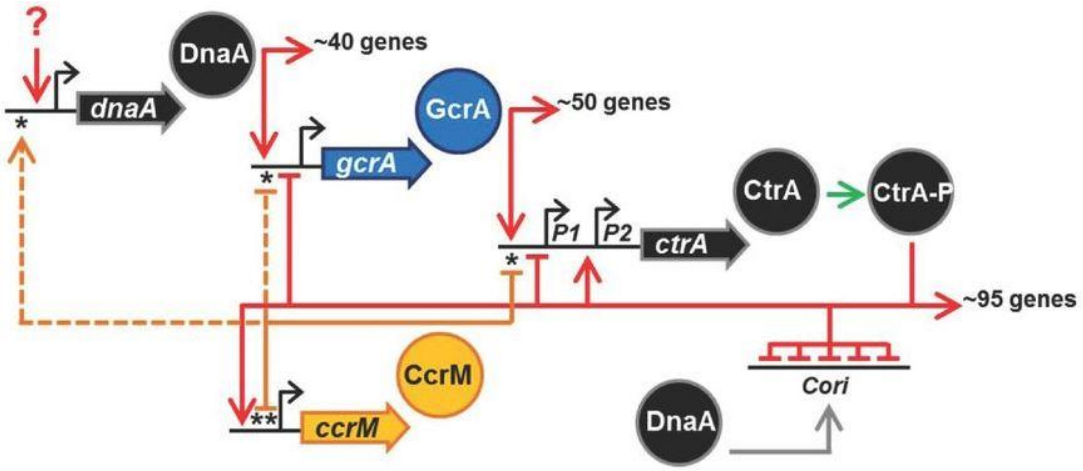


Figure 8. Regulatory circuit of transcription regulators and DNA methyltransferase CcrM. Red arrows indicate transcriptional activation; red dead end signs indicate transcriptional repression and dashed lines indicate regulation under debate. The arrow with question mark represents an unknown regulatory mechanism. Black errors indicate promoters. (Picture taken from (Panis et al. 2014).)

1.3 Microbial hydrogen production

Hydrogen (H_2) is considered to be one of the major energy carriers of the future. This is due to its versatile applications and the lack of environmentally adverse combustion products. H_2 can be burnt with oxygen for heat production, used for electricity generation in fuel cells and has other industrial applications. H_2 has a roughly three times higher energy content per mass than hydrocarbon fuels (Schlapbach and Züttel 2001) and ideal H_2 combustion with oxygen produces only heat and water as final product. Still, it is difficult to store molecular hydrogen due to its chemical and physical properties but constant progress is made in the field (Barthelemy et al. 2016). At the moment, H_2 is produced mainly in an unsustainable manner using fossil fuels (Serban et al. 2003; Dincer and Acar 2015). Other routes for molecular hydrogen production are electrolysis of water (Armaroli and Balzani 2011) or biological hydrogen production (Hallenbeck et al. 2012).

In nature, nitrogenases and hydrogenases are capable of H_2 production (Hallenbeck and Benemann 2002). However, both enzymes classes are generally sensitive to oxygen and the enzymes are either destroyed or inhibited by its presence (Meyer et al. 1978; Colbeau et al. 1980; Vincent et al. 2005; Boyd and Peters 2013). Biotechnologically, current H_2 production attempts include biophotolysis, photofermentation and dark fermentation (Kapdan and Kargi 2006).

In biophotolysis, H_2 is produced by water splitting conducted by cyanobacteria or microalgae. The main problem lies in the necessity for anaerobic conditions for hydrogen production, whereas the capturing of light energy by photosystem II results in oxygen production (Nield et al. 2000). Therefore, a couple of approaches have been applied to reduce oxygen levels, including heavy perfusion of culturing media with inert gas or sulfur depletion in order to reduce photosystem II activity to keep oxygen levels low. However, these approaches all come with the price of low yields (Antal et al. 2003).

In case of photofermentation, purple non-sulfur bacteria use light energy to produce H_2 and carbon dioxide in the process of nitrogen fixation by oxidizing organic acids (Kapdan and Kargi 2006). These bacteria contain only one photosystem and are unable to perform water splitting, omitting the oxygen problem. However, organic substrates are used as a primary energy source and the nitrogen fixation process is

highly energy intensive, making photofermentation not efficient in terms of industrial H_2 production (Hallenbeck and Benemann 2002; Basak and Das 2007).

In dark fermentation processes, organic substrates are oxidized and protons are used as final electron acceptors in order to produce molecular hydrogen by hydrogenases (Kothari et al. 2012). Again, biomass is needed as an energy source, making the process less attractive than direct conversion of sunlight into hydrogen.

The most appealing strategy would be direct H_2 production from water by using sunlight as the main energy source and only having oxygen as side product. Therefore, scientists are actively seeking for enzymes that can catalyze H_2 production and are not sensitive to ambient oxygen levels or new materials that can catalyze hydrogen production from water exploiting solar energy (Wang et al. 2009b; Friedrich et al. 2011).

1.3.1 Classification of hydrogenases

Natural hydrogenases are metalloenzymes that catalyze the reversible oxidation of hydrogen and occur in archaea, bacteria and some eukaryotes (Vignais and Billoud 2007). They can be classified by their cellular localization; there are cytoplasmic hydrogenases that use NADH or NADPH as cofactor, or membrane associated hydrogenases that are coupled to the respiratory electron chain system (Vignais and Billoud 2007). The more common classification system, however, relates to the metal composition of the active site of the hydrogenases. There are [NiFe], [FeFe], and [Fe] hydrogenases with the [NiFe] hydrogenases being the most intensively studied group (Vignais and Billoud 2007). [Fe] hydrogenases are distinct as they need the additional substrate methenyltetrahydromethanopterin and have been found only in methanogenic archaea so far, where they are involved in the conversion of hydrogen and carbon dioxide to methane and water (Thauer et al. 2010).

Under physiological conditions, [FeFe] hydrogenases are predominantly involved in hydrogen production in a strictly anaerobic manner during fermentation processes (Vincent et al. 2007). They occur in bacteria, archaea and some lower eukaryotes and are very sensitive to oxygen and become irreversible inhibited (Vignais and Billoud 2007; Stripp et al. 2009). [NiFe] hydrogenases mainly play a role in hydrogen uptake and hydrogen oxidation (Lubitz et al. 2014). They are found in archaea and bacteria and are generally sensitive to oxygen, however, in contrast to [FeFe] hydrogenases,

inhibition is typically reversible and enzymatic activity can be restored upon reducing conditions (Lenz et al. 2010).

[NiFe] and [FeFe] hydrogenases are phylogenetically unrelated but both comprise a small and a large subunit. The large subunit contains the active site, harboring the diatomic [NiFe] or [FeFe] cluster, respectively. In [NiFe] hydrogenases, the [NiFe] center is coordinated by four cysteine residues and additionally the iron is coordinated by two CN and one CO ligands (Figure 9) (Lubitz et al. 2014). Electrons are transferred to or from the active site via iron sulfur clusters located in the small subunit of the hydrogenases (Figure 9) (Forzi and Sawers 2007).

1.3.2 Genomic organization of hydrogenases and hydrogenase maturation

Hydrogenase genes are organized in operons that contain the genes for the subunits of the hydrogenases, and genes for their maturation. Additional mandatory maturation proteins for [NiFe] hydrogenases are encoded by the *hyp* operon (Blokesch et al. 2002; Forzi and Sawers 2007; Watanabe et al. 2012). The maturation process of hydrogenases includes active site maturation by insertion of the metal ions and CN and CO ligands, proteolytic maturation of the large subunit, insertion of iron sulfur clusters into the small subunit and localization of the hydrogenase complex (Böck et al. 2006).

1.3.3 *Escherichia coli* hydrogenases

In *Escherichia coli* three active [NiFe] hydrogenases have been identified, namely, EcHyd-1, EcHyd-2 and EcHyd-3 (Sawers 1994). EcHyd-1 is reported to be a membrane bound oxygen tolerant hydrogenase. Oxygen tolerance results mainly from an unusual [4Fe-3S] cluster located in the small subunit proximal to the [NiFe] active site in the large subunit. This unusual [4Fe-3S] cluster is coordinated by additional cysteine residues (Fontecilla-Camps et al. 2007; Goris et al. 2011). Oxygen tolerance means, that these hydrogenases can be reactivated quickly after their inactivation by oxygen (Goris et al. 2011) (see also chapter 1.3.4). Closely related membrane bound oxygen tolerant hydrogenases are present in *Ralstonia eutropha* and *Hydrogenvirbio marinus* and show the same mechanism involved in oxygen tolerance (Goris et al. 2011).

EcHyd-2 is an oxygen sensitive membrane bound H_2 uptake hydrogenase. *E. coli* uses it in order to utilize H_2 as an electron donor when no fermentable carbon compounds are available (Lukey et al. 2010). Electrons are being eventually transferred via the quinone pool to fumarate in the process of fumarate respiration (Pinske et al. 2015).

EcHyd-3 is associated with the membrane bound formate-hydrogenlyase complex and contributes to hydrogen production under fermentative conditions (Leonhartsberger et al. 2002).

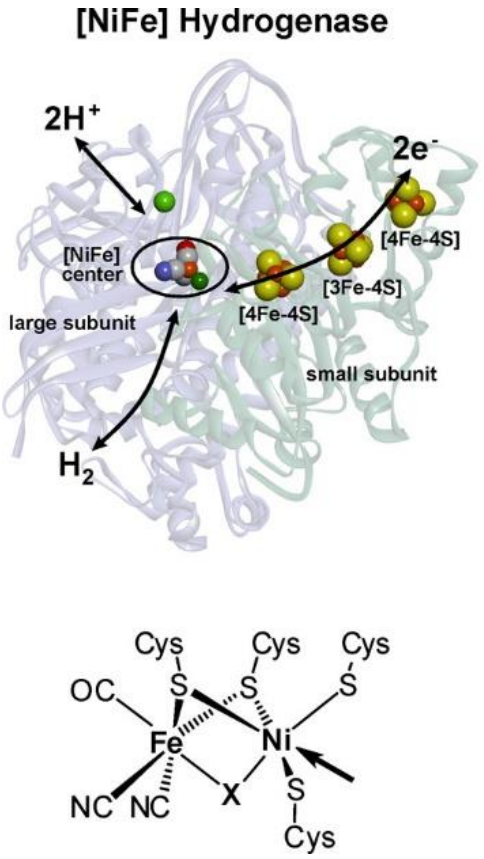


Figure 9. Structure of exemplarily [NiFe] hydrogenase from *Desulfovibrio vulgaris* Miyazaki F with iron sulfur clusters and active site with ligands. (The image is taken from (Lubitz et al. 2014).)

1.3.4 Oxygen tolerant hydrogenases

One key feature of oxygen tolerant hydrogenases is the rapid and complete reduction of oxygen by the transfer of 4 electrons, what subsequently results in water formation. Most described oxygen tolerant hydrogenases contain an unusual [4Fe3S] cluster in the small subunit proximal to the active site (Fritsch et al. 2011; Shomura et al. 2011; Volbeda et al. 2012). In comparison, standard [NiFe] hydrogenases contain a [4Fe4S]

cluster. The [4Fe3S] cluster is coordinated by 6 cysteine residues (instead of 4 in standard hydrogenases) and can store 2 additional electrons in the reduced state that are crucial for rapid oxygen reduction (Lukey et al. 2011).

There are also other mechanisms that might improve oxygen tolerance of hydrogenases. One example are water transfer cavities that aid the release of water molecules after reduction of oxygen (Fritsch et al. 2013). Properties of the gas channel through which H_2 is transported to or away from the active site might also influence oxygen tolerance by precluding oxygen to reach the [NiFe] cluster (Buhrke et al. 2005; Duché et al. 2005). Another mechanism is multimerization of hydrogenases into supercomplexes that exhibit intermolecular electron transfer supporting rapid oxygen reduction in individual oxidized complex partners (Frielingsdorf et al. 2011; Fritsch et al. 2013).

During this work, we identified a bacterial strain that is able to produce considerable amounts of H_2 under aerobic cultivation. The bacterium responsible for hydrogen production belongs to the *Citrobacter* species, and the active hydrogenase was identified to be a [NiFe] type 2 hydrogenase.

2 Principal aims of the study

In the field synthetic biology, scientists use engineering principles to create or modify biological systems. For assembling biological parts in new ways and thereby create systems with new functions or behaviors, the building blocks have to be understood properly to be able to successfully build functional systems.

The cell cycle regulated DNA-(adenine N6)-methyltransferase CcrM from *Caulobacter crescentus* plays a pivotal role in the complex cell cycle of *C. crescentus*. On a biological basis, CcrM is well understood, structurally, however, the available data is limited. CcrM exhibits a C-terminal domain, with unknown function that is conserved in 5'-GANTC-3' targeting homologous enzymes. In this study, it was aimed, to elucidate the potential involvement of the C-terminal domain in DNA binding and target sequence recognition and thereby extend the understanding of the enzyme.

In a second project, it was aimed to use CcrM to artificially and epigenetically regulate gene expression by DNA methylation in a heterologous setup in *Escherichia coli*. Designed zinc finger proteins that bind DNA in a methylation dependent manner were aimed to be used as methylation sensitive transcriptional repressors. It was planned to install a positive feedback system by using CcrM to regulate the expression of its own gene for the construction of a system that would feature synthetic epigenetic memory in *E. coli*.

Synthetic biology approaches are also applied in the field of sustainable energy production. In previous work, a bacterium was discovered that produced molecular hydrogen under aerobic cultivation conditions. In the scope of this thesis, it was aimed to identify this bacterium, to determine the responsible hydrogen producing enzyme, and to investigate its potential applications.

3 Results

One major aim of synthetic biology is the development of new biological functions and their application in biotechnological or medical settings. In order to be able to construct new systems, it is of great importance to describe and understand the behavior and characteristics of the used parts and modules, be it functional modules or regulatory systems.

With the aim of developing the necessary understanding of potential biological components for synthetic circuit construction, the bacterial DNA-(adenine N6)-methyltransferase CcrM from *Caulobacter crescentus* was investigated structurally and biochemically (3.1). Using CcrM in a heterologous set up together with an engineered zinc finger protein, epigenetic regulatory systems with memory function have been designed and developed. The developed system was coupled to different sensor devices, and it was shown that the systems is capable of recording certain transient stimuli and memorize these events even after cessation of the trigger signal for many bacterial generations (3.2).

Additionally, an unusual oxygen tolerant hydrogenase was isolated and identified from a *Citrobacter* species. Heterologous expression and activity measurements suggest potential applications of this novel enzyme as a building block for synthetic biology applications in the field of industrial hydrogen production (3.3).

3.1 Investigation of the cell cycle regulated DNA-(adenine N6)-methyltransferase from *Caulobacter crescentus*

The cell cycle regulated DNA-(adenine N6)-methyltransferase plays a critical role in the complex cell cycle of *Caulobacter crescentus*. The biological role of CcrM is well studied (Gonzalez et al. 2014) and there are several biochemical reports on CcrM (Shier et al. 2001; Albu et al. 2012a; Albu et al. 2012b). However, structural information of the enzyme is limited. Interestingly, sequence alignments of CcrM from *Caulobacter* and homologs revealed a conserved C-terminal domain with unknown function (Figure 1 B from Appendix 1). During this study, the C-terminal domain of *Caulobacter crescentus* CcrM was investigated and showed an involvement in target sequence recognition. Thus a reclassification of CcrM into the δ-class of DNA-(adenine N6)-

methyltransferases has been proposed. The results below have been published in *Biochimie* and will be only briefly summarized here. For more details, see (Maier et al. 2015a) in Appendix 1.

3.1.1 The CTD of CcrM is essential for catalytic activity

The primary protein structure of CcrM from *Caulobacter crescentus* and homologous DNA-(adenine N6)-methyltransferases that show specificity for 5'-GANTC-3' target sites were analyzed. By applying a multiple sequence alignment, a conserved C-terminal domain was identified, only present in 5'-GANTC-3' specific prokaryotic DNA MTases (Supplementary Figure 1 from Appendix 1). Deletion of the C-terminal domain of *Caulobacter crescentus* CcrM led to loss of its catalytic activity and loss of DNA binding. The C-terminal domain alone did also not show DNA binding (Figure 4 from Appendix 1). Next, 13 conserved amino acids in the C-terminal domain of *Caulobacter crescentus* CcrM were mutated to alanine (Figure 1 from Appendix 1). Individual mutant proteins were heterologously expressed in *E. coli* cells and purified by affinity chromatography (Figure 2A from Appendix 1). Catalytic activities of the individual mutant proteins were determined on a 23 bp DNA substrate with one hemi-methylated 5'-GANTC-3' site. Six of the 13 mutant proteins showed wild type-like activity and seven mutant proteins showed reduced methylation activity (Supplementary Figure 4 from Appendix 1). In order to rule out secondary structural changes as cause for the reduced catalytic activity, circular dichroism measurements were employed and no significant differences between wild type and mutant proteins were found (Figure 4D from Appendix 1).

3.1.2 CTD of CcrM is not involved in AdoMet binding

In DNA-(adenine N6)-MTases Sequence Motifs I, II, III, and X form the AdoMet binding site. These motifs can be easily identified by characteristic conserved amino acid sequences and in CcrM, these are not located in the C-terminal domain (Supplementary Figure 1 from Appendix 1) (Cheng 1995; Malone et al. 1995; Jeltsch 2002). In order to preclude an involvement of the C-terminal domain in cofactor (AdoMet) binding experimentally nevertheless, an AdoMet binding assay was performed and AdoMet binding of CcrM wild type was compared with the mutant

proteins that showed decreased catalytic activity. However, AdoMet binding was not impaired significantly in any of the mutant proteins (Figure 3 from Appendix 1).

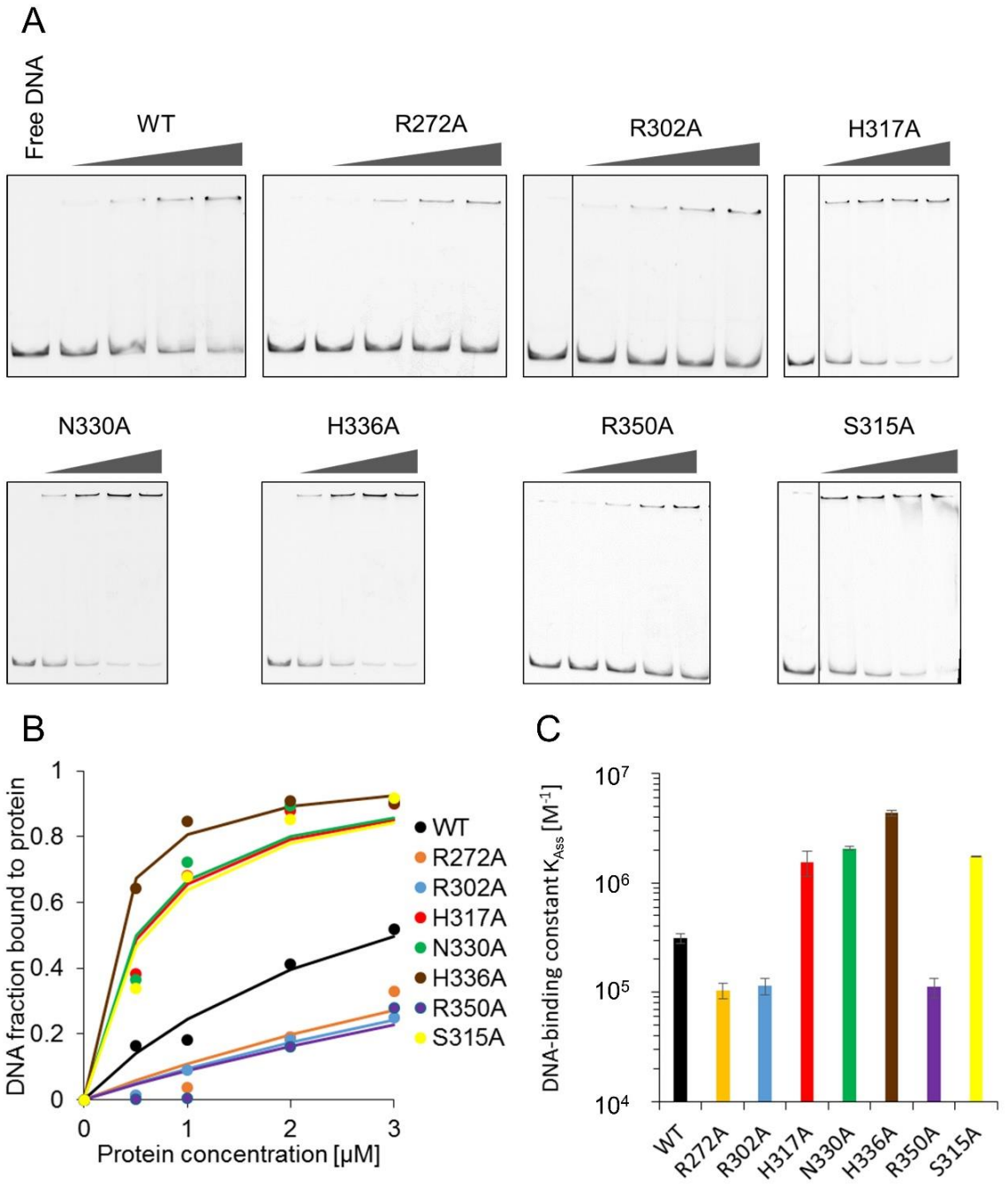


Figure 10. DNA binding of CcrM wild type and mutants. A) Examples of EMSA experiments, which were carried out using 15 nM DNA (a Cy-5 labelled 152-base pair PCR product) and varying CcrM concentrations (0.5 µM, 1 µM, 2 µM, and 3 µM). B) Examples of titration curves obtained after densitometric analyses of EMSA gels with CcrM wild type and mutants. C) Calculated K_{Ass} values of different CcrM mutants. Error bars indicate standard deviations of at least two independent experiments. (The image and figure legend are taken from Appendix 1)

3.1.3 C-terminal domain of CcrM is involved in DNA binding

Next, the DNA binding behavior of the CcrM mutant proteins with decreased catalytic activity was studies by using an EMSA assay. Interestingly, several of the mutant proteins showed altered DNA binding behavior compared to wild type CcrM protein (Figure 10). Mutants R272A, R302A and R350A showed approximately three times decreased DNA binding and mutants S315A, H317A, N330A, and H336A showed increased DNA binding in the range of five to 14 times. In conclusion, the experimental data reveal that the C-terminal domain of *Caulobacter crescentus* is involved in DNA binding. The consequences of this finding will be discussed below.

3.2 Design of synthetic epigenetic circuits

Utilizing the CcrM methyltransferase and a designed zinc finger (ZnF) protein repressor, a synthetic epigenetic memory system was developed in an iterative design process in *Escherichia coli*. The system is able to store transient sensory information in form of stably maintained DNA methylation patterns in living bacteria for many cell generations. The results shown here are only a succinct summary. For a more detailed description of all experiments, refer to (Maier et al. 2017) in Appendix 2.

At first, ZnF proteins were designed to bind DNA sensitive to DNA-(adenine N6)-methylation and analyzed with a bacterial two-hybrid system (Supplementary Figure 2, 3, 4, 5 from Appendix 2) (Wright et al. 2006). In the next step, one of these engineered ZnF proteins was used and repurposed to act as a transcriptional repressor and a promoter/repressor system that can be induced by adenine-N6 methylation in 5'-GANTC-3' sites was created (Supplementary Figure 6 from Appendix 2).

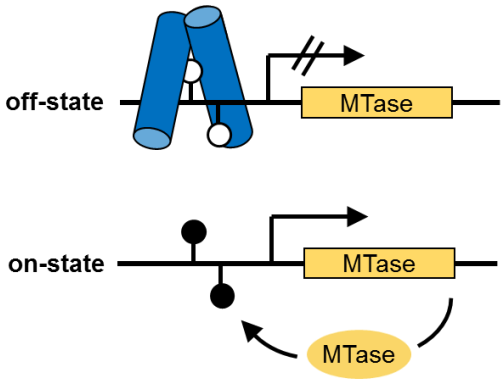


Figure 11. Scheme of the synthetic epigenetic system featuring positive feedback. In the off-state, the synthetic ZnF repressor binds the promoter region of a methyltransferase (MTase) gene. In the on-state, the promoter region is methylated, repressor binding is hindered and the MTase gene is transcribed resulting in a positive feedback loop. Filled and open lollipops represent methylated and unmethylated 5'-GANTC-3' sites. (The image is taken from Appendix 2. The figure legend is modified from Appendix 2.)

Next, this methylation sensitive promoter repressor system was used to regulate the expression of a *ccrM* gene and thereby an epigenetic system with positive feedback was created (Figure 11). This system exhibits two stable states, an off-state and an on-state (Figure 12, Figure 1 from Appendix 2, Supplementary Figure 8 from Appendix 2). The off-state is maintained by the engineered ZnF repressor and the on-state by constant promoter methylation by CcrM, which hinders ZnF DNA binding. An *egfp*

(enhanced green fluorescent protein) gene was cloned as a reporter in front of the *ccrM* gene and this set up was called reporter-maintenance operon.

The design process of the epigenetic memory system included multimerization of the ZnF protein and the introduction of multiple ZnF binding sites in the promoter region of the system in order to reach a stable off-state (Supplementary Figure 6 and 8 from Appendix 2). Autoregulation of the ZnF repressor was shown to be necessary for a low and constant expression, allowing a stable on-state (Supplementary Figure 10 from Appendix 2).

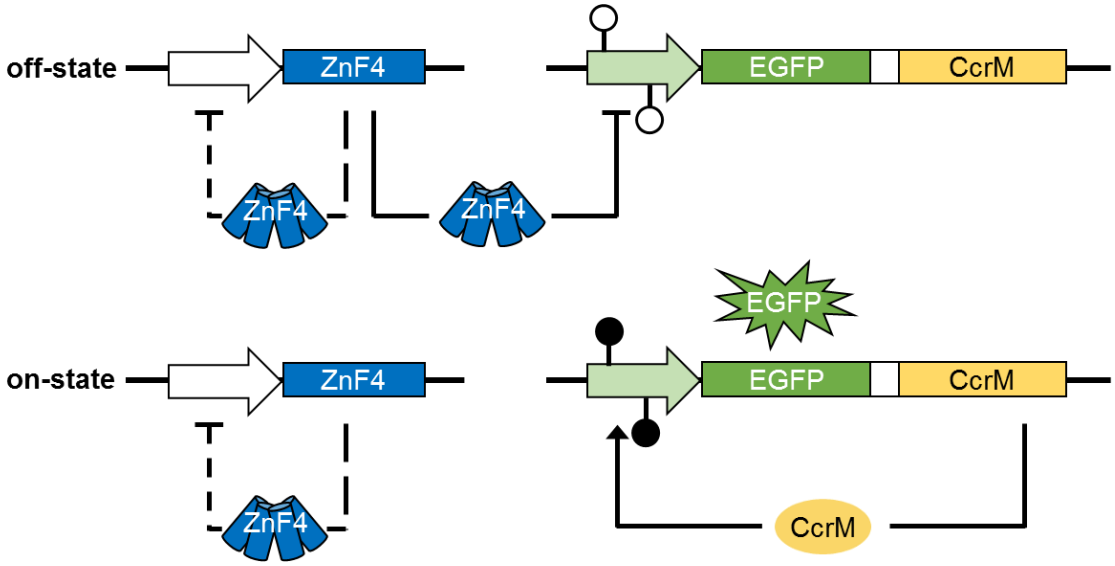


Figure 12. Schematic circuit design of the synthetic epigenetic memory system with ZnF repressor (ZnF4) and reporter-maintenance operon, consisting of an *egfp* gene and a *ccrM* gene. In the off-state, the ZnF repressor inhibits transcription of the reporter-maintenance operon by binding to the unmelthylated promoter region (empty lollipops). Once the system is switched to the on-state, EGFP and CcrM are expressed and binding of the ZnF repressor to the promoter region of the reporter-maintenance operon is prevented by methylation (filled lollipops). CcrM constantly re-methylates the operator ZnF binding sites and keeps the system in the on-state. The ZnF repressor regulates its own expression in a manner not affected by methylation (indicated negative feedback with dashed lines). (The image is adapted from Appendix 2, the figure legend is modified from Appendix 2.)

Upon different trigger signals, the expression of the reporter-maintenance operon can be initiated, either by loss of the ZnF protein binding due to thermal instability or by introduction of initial DNA methylation at the ZnF binding site that interferes with ZnF protein binding. Subsequently, the maintenance CcrM methyltransferase methylates the promoter region of the reporter-maintenance operon and keeps the system in the

on-state (Figure 1 D from Appendix 2). We coupled the input signals heat, nutrients (arabinose), and DNA damage (ultraviolet (UV) light or cisplatin) to the reporter-maintenance module to switch from the off- to the on-state (Figure 12, Figure 1 and 3 from Appendix 2).

For switching the system by arabinose, a trigger operon was used that consists of a gene coding for CcrM (trigger CcrM) and a *mCherry* reporter gene under the control of a pBAD promoter (Figure 12A). Upon induction of the system with arabinose, the trigger CcrM and mCherry are expressed. The trigger CcrM can methylate the promoter of the reporter-maintenance operon and hence switch the memory system to the on-state. Acute induction of the system could be observed by tracking the mCherry fluorescence signal (Figure 12B). After removal of arabinose from the medium, mCherry fluorescence went back to background levels. EGFP levels increased during induction of the system to on-state levels and after removal of arabinose remained in this state for at least 96 h (roughly 48 cell generations) (Figure 12C). As a negative control, the same system was cloned with an inactive maintenance CcrM. With this control system, on-state EGFP levels could be only observed during acute induction of the system with arabinose. After removal of arabinose, EGFP fluorescence levels went back to background levels indicating that no stable on-state was reached (Figure 12C). Additionally to the standard analysis of fluorescence intensities by fluorescence spectroscopy, cells in the on- and off-state were analyzed by flow cytometry and confocal laser scanning microscopy (Figure 12D, 12E, and 12F).

Exploiting the epigenetic nature of the memory system, a reset function was implemented in order to switch back from the on- to the off-state. To achieve this back-switching, the maintenance CcrM was fused with an *mf*-Lon specific degradation tag, allowing a selective and inducible degradation of the methyltransferase and hence enforced a reset of the system to the off-state (Figure 4 from Appendix 2, Supplementary Figure 19 and 20 from Appendix 2).

In summary, synthetic epigenetic circuits that can sense different transient stimuli and store this information in form of DNA methylation patterns for many cellular generations in *Escherichia coli* have been constructed. Further, the epigenetic nature of the developed system was exploited and it was shown that the system is resettable. Potential applications and extensions of the system are discussed below.

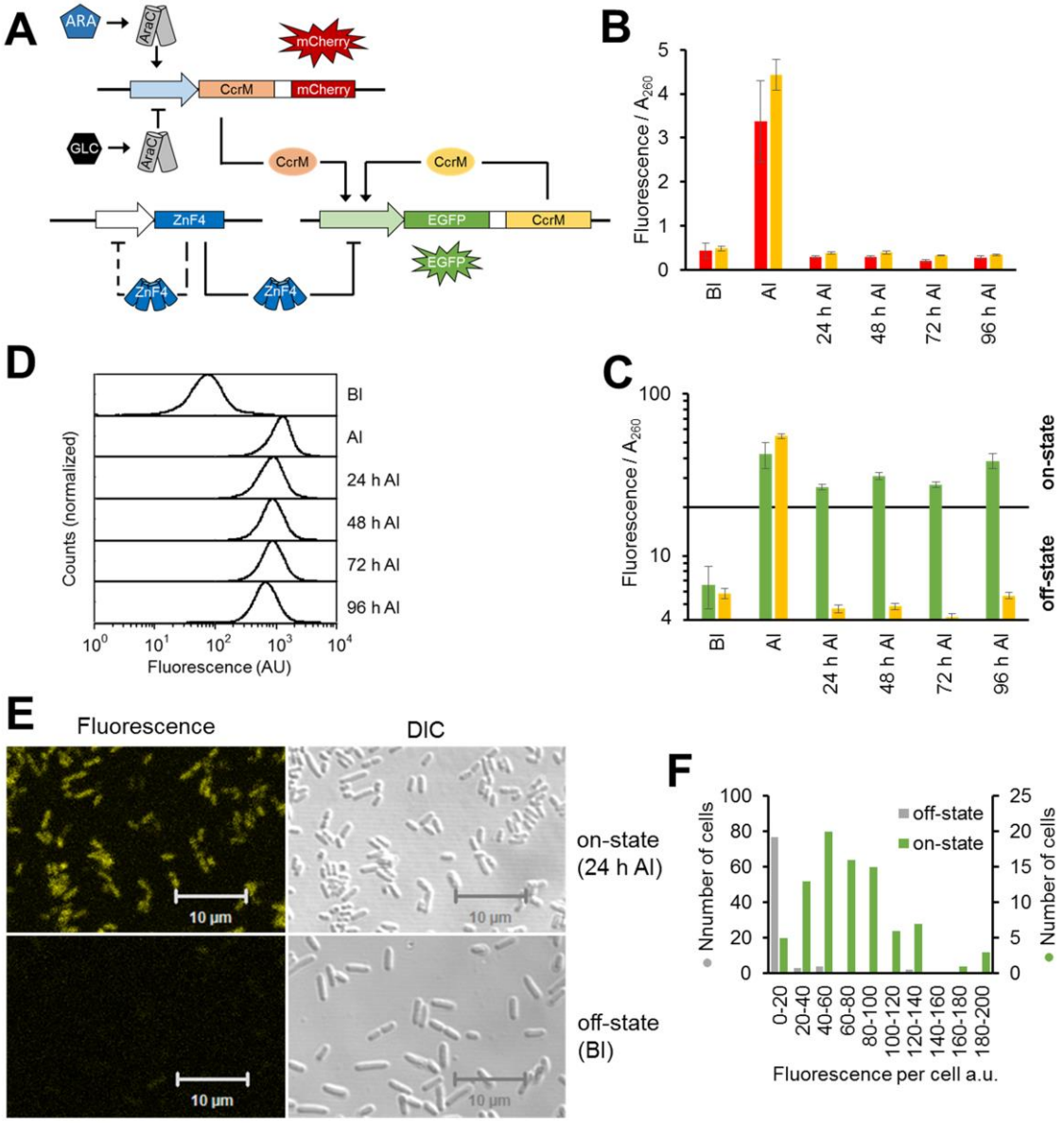


Figure 13. Chemical induction of memory system by arabinose. A) Circuit design of memory system. In the initial off-state, glucose represses the trigger CcrM expression. The on-state can be induced by arabinose supplementation, which induces CcrM and mCherry expression from the trigger plasmid. After removal of arabinose, the on-state is maintained by expression of the reporter-maintenance operon via a positive feedback loop. B) Red fluorescent protein (mCherry) fluorescence measured in total cell lysate. Red bars: WT CcrM in the reporter-maintenance operon; orange bars: active site mutant CcrM in the reporter-maintenance operon. (BI: before induction, i.e. in glucose containing media; AI: after induction, i.e. cultivation for 12 h in presence of arabinose; 24 h AI, 48 h AI, 72 h AI, 96 h AI, i.e. cultivation in glucose containing media for the indicated time period after cultivation for 12 h in arabinose containing media) (error bars indicate SD, n=3). mCherry signal can only be detected in the presence of arabinose. C) EGFP signal measured in total cell lysate. Upon induction with arabinose, green fluorescence levels rise above the threshold level and stay in the on-state range for at least 96 h after induction (green bars). If the catalytically inactive

CcrM is present in the reporter-maintenance operon, EGFP fluorescence disappears 24 h after induction and no memory function can be observed (orange bars) (error bars indicate SD, n=3). D) EGFP signal measured by flow cytometry. Cells with functional reporter-maintenance operon (i.e. wildtype CcrM) were analysed. Histograms depict homogeneous populations in the off- and on-state. 20,000 events were collected for each measurement. E) Confocal laser scanning microscopy pictures and differential interference contrast (DIC) microscopy pictures of cells with functional reporter-maintenance operon in the on-state (24 h AI) and cells in the off-state (BI). F) Quantification of fluorescence intensities of cells in the on-state and in the off-state recorded by confocal laser scanning microscopy (n=87, per state). (The image and the figure legend (modified) are taken from the manuscript attached as Appendix 2.)

3.3 Identification, cloning and heterologous expression of *Citrobacter* sp. SG hydrogenase 2

Molecular hydrogen (H_2) can be used as a pollution free energy carrier, as the combustion products under ideal conditions are only water and heat. However, today, H_2 is mainly produced by steam reforming of hydrocarbons (Dincer and Acar 2015). Biological hydrogen production would be more environmentally friendly and sustainable. Ideally, biological hydrogen production would be performed by photoautotrophic organisms without the need of intermediate accumulation of biomass. Known enzymes capable of hydrogen production (hydrogenases) are very sensitive to molecular oxygen what makes it difficult to work with in biotechnological and synthetic biology approaches. Therefore, the field is looking for oxygen tolerant enzymes that would ease applications and development of synthetic systems for biological hydrogen production. Additionally, hydrogen detection demands for high-end instruments like gas chromatographs. Here, an easy to use and quantitatively operating semiconducting device for the reliable detection and measurement of molecular hydrogen in the gas phase was introduced. Using this device, a bacterial strain was identified, naturally producing considerable amounts of hydrogen under aerobic conditions. The responsible enzyme was identified and it was shown that it is also active when heterologously expressed in *Escherichia coli*. The results below have been published in *Journal of Biotechnology* and will be only briefly summarized. For more details, refer to (Maier et al. 2015b) in Appendix 3.

3.3.1 Easy-to-use hydrogen detector

A semi-conducting device (Hydrogen Leak Detector H2000 (Sensistor)) was used for the detection of molecular hydrogen, produced by microorganisms and the specificity of the detection system was confirmed by gas chromatography (Figure 1 from Appendix 3). The dynamic range of the device ranges from below 10 parts per million (ppm) up to 4-digit ppm values.

3.3.2 Identification of hydrogen producing bacterium *Citrobacter* sp. SG and the active hydrogenase

A hydrogen producing bacterium that generates considerable amounts of molecular hydrogen during aerobic culturing was discovered. Identification by 16S rDNA sequencing revealed that a close relationship to *Citrobacter freundii* and the new isolate was called *Citrobacter* sp. SG (for Stuttgart Germany) (Figure 3A from Appendix 3, GenBank: AF025365.1). Hydrogen production of *Citrobacter* sp. SG from glucose was analyzed under multiple conditions including a 1 L culture with a constant airflow of 75 mL/min (Figure 2 from Appendix 3). Next, the active hydrogen producing enzyme was identified by consecutively applying different protein purification steps. A 1 liter overnight shaking flask culture of *Citrobacter* sp. SG was grown in a 5 L baffled flask and was used as starting material for the protein enrichment. Cells were pelleted and lysed by sonication, the membrane fraction was isolated and three chromatographic protein purification steps were applied (mixed mode ion exchange, strong anion exchange, gelfiltration). Hydrogenase activity could be enriched approximately 120 fold. Hydrogenase containing fractions were analyzed by SDS-PAGE and dominant visible protein bands were applied to mass spectroscopy. Small and large subunits of hydrogenase 2 could be detected by peptide mass fingerprinting (Figure 3B from Appendix 3).

3.3.3 Cloning and heterologous expression of *Citrobacter* sp. SG [NiFe]-hydrogenase 2

The hydrogenase 2 operon from *Citrobacter* sp. SG was sequenced (GenBank: KP704659) and cloned into an inducible pET28a (+) vector for heterologous overexpression in *Escherichia coli* BL21-CodonPlus™ (DE3) cells. Protein expression via induction by IPTG supplementation showed hydrogenase activity in a Ni^{2+} dependent manner (Figure 4 from Appendix 3).

3.3.4 Multiple sequence alignment of hydrogenase 2 large and small subuits

There have been recent reports of a *Citrobacter* species (*Citrobacter* sp. S-77) that expresses an oxygen tolerant membrane bound hydrogenase (Eguchi et al. 2012; Matsumoto et al. 2014). Later, it turned out that the enzyme identified in these papers

is also a type 2 hydrogenase (Muhd Noor et al. 2016). Multiple sequence alignments of small and large subunits of hydrogenase 2 from *Citrobacter* sp.77, *Citrobacter freundii*, *Citrobacter* sp. SG and *Escherichia coli* and hydrogenase 1 form *Escherichia coli* revealed that sequences of all three *Citrobacter* species are highly similar, at least in the primary protein structure (Appendix 4 and 5).

There are crystal structures available for the oxygen tolerant hydrogenase 1 from *Escherichia coli* and the membrane bound [NiFe] hydrogenase from *Ralstonia eutropha*. These data suggest that the proximal unusual [4Fe3S] cluster present in these enzymes is responsible for their oxygen tolerance (Volbeda et al. 2012; Frielingsdorf et al. 2014; Evans et al. 2016). The unusual [4Fe3S] cluster is coordinated by two additional cysteine residues in comparison to standard membrane bound hydrogenases which carry a [4Fe4S] proximal cluster (Figure 14). These two additional cysteine residues are not present in the small subunit of type 2 hydrogenases from either *Escherichia coli* or *Citrobacter* sp. (Appendix 4). Interestingly, there are two additional cysteine residues present in the large subunit of the *Citrobacter* sp. hydrogenases 2, which are absent in *Escherichia coli* hydrogens 2. However, alignments of amino acid sequences and structural comparison reveled, that the 2 additional cysteine residues in the *Citrobacter* sp. hydrogenases 2 are not in spacial proximity to the iron sulfur clusters of the small subunit. According to a comparison with the crystal structure of hydrogenase 1 from *Escherichia coli,* they are rather located at the surface of the hydrogenase large subunits (Figure 14).

In conclusion, a hydrogen producing bacterium was identified as a *Citrobacter* sp. and the active enzyme to be a type 2 hydrogenase. It was also possible to actively express the hydrogenase 2 from *Citrobacter* sp. SG in *Escherichia coli*. The detailed mechanism of its oxygen tolerance is yet unknown.

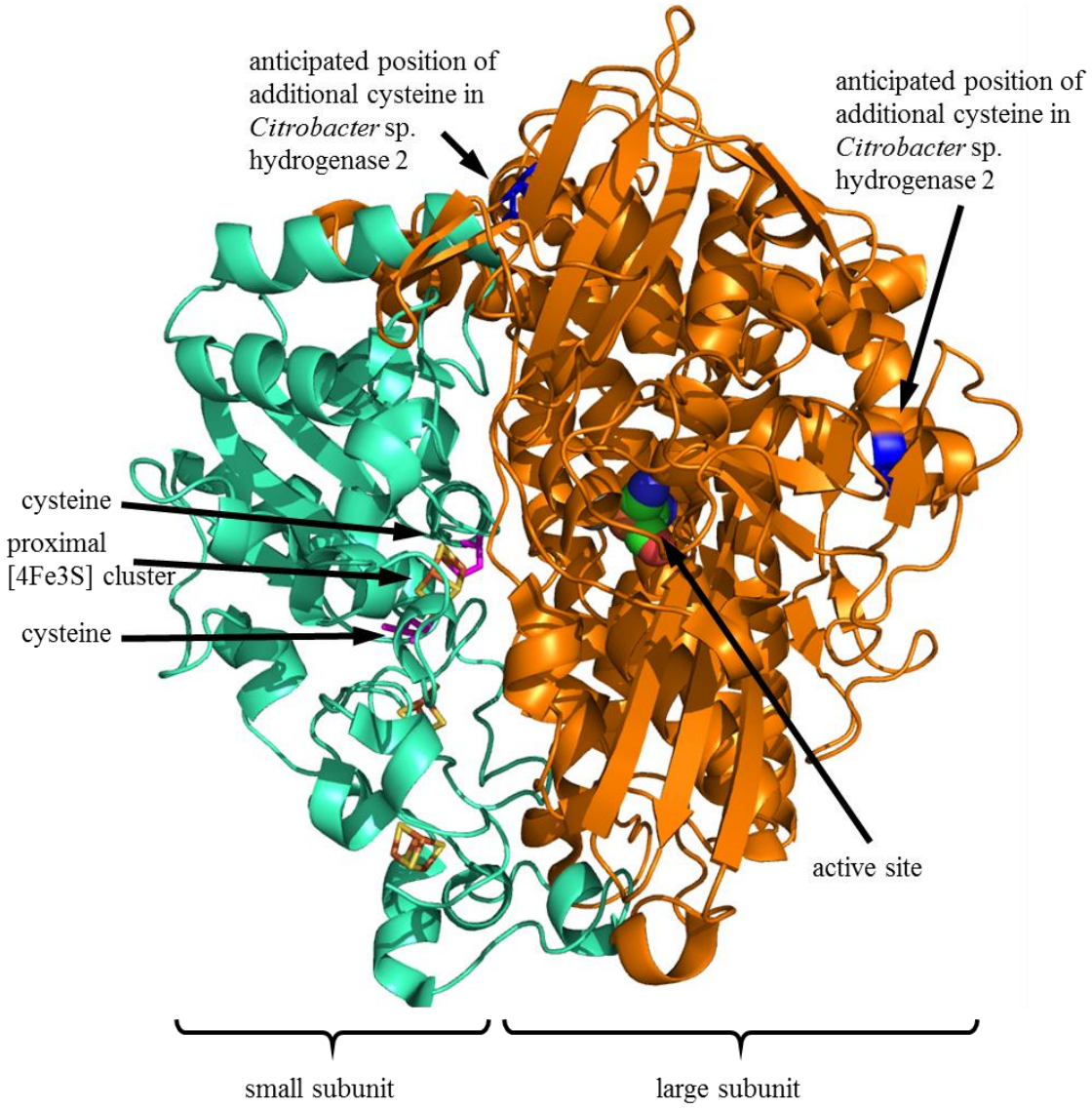


Figure 14. Crystal structure of small (cyan) and large (orange) subunit of the oxygen tolerant hydrogenase 1 from *Escherichia coli* in cartoon representation. (PDB ID: 5A4F (Evans et al. 2016)) [NiFe] active site with CN and CO ligands in sphere representation. The three iron sulfur clusters in the small subunit responsible for electron transfer to, or away from the active site are represented as sticks. Characteristic cysteine residues for oxygen tolerant hydrogenases, necessary for the coordination of the proximal [4Fe3S] cluster are indicated as magenta sticks. Anticipated positions of additional cysteine residues present in the large subunit of hydrogenase 2 of *Citrobacter* sp. are indicated as blue sticks. Their position indicates no proximity to the iron sulfur clusters of the small subunit.

4 Materials and Methods

4.1 Investigation of the cell cycle regulated DNA-(adenine N6)-methyltransferase from *Caulobacter crescentus*

4.1.1 Multiple sequence alignment

Multiple sequence alignment of DNA-(adenine N6)-methyltransferases targeting 5'-GANTC-3' sites and other β-class DNA-(adenine N6)-methyltransferases was generated with ClustalW (Thompson et al. 2002) using default parameter settings. For visualization, the BioEdit software was used (Hall 1999).

4.1.2 Site-directed mutagenesis, heterologous protein expression and protein purification

Site-directed mutagenesis was conducted by PCR based megaprimer method (Jeltsch and Lanio 2002). CcrM proteins were overexpressed in *E. coli* ER2566 (NEB) cells and purified by affinity chromatography as described in more detail in the methods section of the manuscript, which is attached as Appendix 1.

4.1.3 Methylation activity assay, secondary structure analysis and AdoMet binding

For catalytic activity measurements of CcrM mutant proteins, a biotin-avidin microplate assay (Roth and Jeltsch 2000) was used that deploys ^{3}H-labeled AdoMet and the transfer of radioactively labeled methyl groups is measured. Potential secondary structure differences between mutant proteins and wild type CcrM were investigated by circular dichroism spectrophotometry. Binding of AdoMet by mutant proteins and wild type CcrM was examined by a fluorescence based competition assay (Schluckebier et al. 1997; Jurkowski et al. 2012). For a more detailed description of the methods mentioned above, refer to the method section in the manuscript attached as Appendix 1.

4.1.4 DNA binding assays

DNA binding of N-terminal and C-terminal domains of *Caulobacter crescentus* CcrM was investigated by a radioactive filter binding assay. DNA binding of CcrM wild type and mutant proteins was studied with an electrophoretic mobility shift assay. Both methods are described in more detail in the methods section in the attached manuscript in Appendix 1.

4.2 Design of synthetic epigenetic circuits

Escherichia coli XL-1 Blue cells were used to test all designed synthetic epigenetic constructs and circuits. Cells were grown at 30°C in test tubes with a culture volume of 5 mL, unless stated otherwise in the more detailed description of the methods in the attached manuscript in Appendix 2.

4.2.1 Molecular biology work

Gibson assembly was used for all cloning procedures (Gibson et al. 2009; Gibson et al. 2010b). Q5® High-Fidelity DNA Polymerase was used for PCRs (NEB). For more details and exceptions, refer to the methods in the manuscript attached in Appendix 2.

4.2.2 Zinc finger binding assays

4.2.2.1 Zinc finger array DNA binding assay

Zinc finger proteins consisting of an array of three concatenated zinc finger domains were tested for binding to their anticipated 9 bp target sites by using a bacterial two-hybrid reporter system (Wright et al. 2006). In this assay DNA binding is indicated via the expression of a reporter gene (*lacZ*). The zinc finger protein is fused to Gal11P, which can interact with Gal4, which in turn is fused to RNA polymerase. As the anticipated DNA binding site is placed upstream of the lacZ gene, DNA binding of the zinc finger protein triggers transcription of the *lacZ* reporter gene. The assay was modified to test DNA binding of methylated binding sites, by coexpression of CcrM. For a more detailed description of the assay, see the attached manuscript in Appendix 2.

4.2.2.2 ZnF4 binding assay

DNA binding preference for unmethylated DNA compared to methylated DNA of the engineered zinc finger protein used as a methylation sensitive transcriptional repressor (ZnF4) was analyzed by an electromobility shift assay (EMSA). For more details, refer to the methods described in Appendix 2.

4.2.3 Cloning, overexpression and purification of ZnF4

ZnF4 was cloned in pMAL-c2X (NEB) vector resulting in a MBP-ZnF4 fusion protein. *E. coli* BL21-CodonPlus™ (DE3) cells were used for overexpression of the fusion protein. Purification of the ZnF4 fusion protein was performed by affinity chromatography (amylose resin). For more details, refer to the methods described in Appendix 2.

4.2.4 Fluorescence measurements

4.2.4.1 Fluorescence measurements in cell lysates

For measurements of the expression of reporter genes *egfp* or *mCherry*, fluorescence was measured in cell lysates with a Jasco FP-8300 Fluorescence Spectrometer. For more details, including cell lysis and spectrometer settings, see the method section in the attached manuscript in Appendix 2.

4.2.4.2 Microscopic fluorescence imaging

Microscopic pictures of whole cells were recorded with a LSM 710 Zeiss confocal microscope. Fluorescence intensities were quantified with an image processing program (ImageJ). For more details, see the method section in the attached manuscript in Appendix 2.

4.2.4.3 Flow cytometry measurements of EGFP expression

Flow cytometry was used as an additional method to analyze fluorescence intensities of whole cells. A BD FACS Calibur flow cytometer was used and the obtained data was analyzed using FCS Express V4 (De Novo Software). For details of sample preparation and measurement settings, see the method section in the attached manuscript in Appendix 2.

4.2.5 Promoter methylation analysis

To assign the methylation status of the promoter region of the epigenetic memory system, an assay was developed, that is based on methylation sensitive restriction digest followed by quantification of undigested DNA by qPCR. A CFX96 Connect Real-Time detection system (Bio-Rad) and SsoFast EvaGreen supermix (Bio-Rad) was used. For a detailed method description, see the method section the attached manuscript in Appendix 2.

4.3 Identification, cloning and heterologous expression of *Citrobacter* sp. SG hydrogenase 2

4.3.1 Hydrogen measurements

Hydrogen measurements were performed using a semiconducting device (Hydrogen Leak Detector H2000, Sensistor). In order to validate the device and exclude cross reactivity, selected samples were analyzed by gas chromatography (Agilent 7890 gas chromatography system). Hydrogenase activity in cell lysates and partially purified extracts was detected via a methyl viologen (MV) / sodium dithionite assay using the semiconducting device for detection of evolved hydrogen. For more details of the applied methods, refer to the manuscript, which is attached as Appendix 3.

4.3.2 Bacterial culturing

Citrobacter sp. SG was cultured in shaking flaks at ambient oxygen conditions at 30°C in LB medium supplemented with a mix of various metal ions (described in detail in the manuscript in Appendix 3). For hydrogen production measurements with air perfusion, overnight cultures were cooled from 30°C to room temperature, transferred to 2 liter round bottom flasks and flushed with air. *Citrobacter* sp. SG and *Escherichia coli* BL21-CodonPlus™ (DE3) were treated in the exact same way. Recombinant *Escherichia coli* BL21-CodonPlus™ (DE3) inducibly expressing *Citrobacter* sp. SG hydrogenase 2 were cultured at 18°C for protein expression and at room temperature for hydrogen production measurements. Recombinant *E. coli* cells grown in LB medium supplemented with the metal ion mix. For more details of the applied methods, refer to the manuscript, which is attached as Appendix 3.

4.3.3 Genomic DNA preparation and 16S rDNA sequencing

For identification of the hydrogen producing bacterium (later identified as a *Citrobacter* sp. and then called *Citrobacter* sp. SG for Stuttgart, Germany), genomic DNA was isolated and 16S rDNA sequencing was performed. To isolate the genomic DNA the proteinase K/sodium dodecyl sulfate/phenol extraction method was used (Herrmann and Frischauf 1987). 16S rDNA was amplified by PCR, cloned by TA cloning (StrataClone PCR Cloning Kit, Agilent) and sequenced. For more details, primer sequences and GenBank entries, refer to the manuscript, which is attached as Appendix 3.

4.3.4 Hydrogenase enrichment and peptide mass fingerprinting

For identification of the active enzyme responsible for hydrogen production in *Citrobacter* sp. SG, a 1 L overnight culture of *Citrobacter* sp. SG grown at 30°C in a 5 L baffled flask under ambient oxygen levels was used as starting material. After cell lysis, the membrane fraction was isolated and solubilized. Chromatographic purification steps (hydroxyapatite, anion exchange, gelfiltration) were applied. For selection of fractions used in subsequent enrichment steps, the methyl viologen/ sodium dithionite assay was used. Hydrogenase containing fractions were separated by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and dominant protein bands were analyzed by peptide mass fingerprinting. For a more detailed description of the hydrogenase enrichment procedure and the identification of the active hydrogenase, refer to the manuscript, which is attached as Appendix 3.

4.3.5 Cloning of hydrogenase 2 operon of *Citrobacter* sp. SG

The operon coding for hydrogenase 2 from *Citrobacter* sp. SG was amplified by PCR from isolated genomic DNA and cloned into an inducible pET-28(+) vector. For GenBank entry of the DNA sequence of hydrogenase 2 from *Citrobacter* sp. SG and a more detailed cloning description, refer to refer to the manuscript, which is attached as Appendix 3.

4.3.6 Hydrogenase sequence alignment

Sequences for amino acid alignments of hydrogenases shown in Appendix 4 and 5:

Large subunits

GenBank: AOM69335.1 (Hydrogenase 1 large subunit, *Escherichia coli*)
GenBank: OJF21604.1 (Hydrogenase 2 large subunit, *Escherichia coli*)
GenBank: OIY06664.1 (Hydrogenase 2 large subunit, *Citrobacter freundii*)
GenBank: AKJ80129.1 (Hydrogenase 2 large subunit, *Citrobacter* sp. SG)
GenBank: GAN52682.1 (hydrogenase 2 large subunit, *Citrobacter* sp. S-77)

Small subunits

GenBank: OJF21607.1 (Hydrogenase 1 small subunit, *Escherichia coli*)
GenBank: SCQ06936.1 (Hydrogenase 2 small subunit, *Escherichia coli*)
GenBank: ALD78266.1 (Hydrogenase 2 small subunit, *Citrobacter freundii*)
GenBank: AKJ80127.1 (Hydrogenase 2 small subunit, *Citrobacter* sp. SG)
GenBank: GAN52685.1 (Hydrogenase 2 small subunit, *Citrobacter* sp. S-77)

Sequence alignments were performed using ClustalW with standard settings (Thompson et al. 2002) and visualized using jalview (Waterhouse et al. 2009) using clustalX coloring.

5 Discussion

The work of synthetic biologists resembles partly the work of engineers. Synthetic biologists use and assemble biological parts aiming to create new functionality. To be effective, it is of great importance to know basic design principles in natural systems and to understand the used parts as much as possible. With this knowledge, synthetic biologists can program cells or organisms to conduct a large variety of different functions or to produce desired compounds. Synthetic biology approaches find their way into biotechnological industries, fuel production, agriculture, environmental monitoring, waste decontamination and bioremediation, healthcare, biocomputing and many more fields (Khalil and Collins 2010; Smanski et al. 2016). Additionally, in the long run, synthetic biologist are trying to create fully synthetic life (Gibson et al. 2010a; Hutchison et al. 2016).

In this thesis, scientific work has been conducted that helps to better understand biological parts; referring to the *Caulobacter crescentus* cell cycle regulated methyltransferase and the unusual *Citrobacter* sp. hydrogenase 2. Furthermore, knowledge on epigenetics, DNA methyltransferases, bacterial gene regulation and protein engineering has been applied to build novel synthetic epigenetic devices in bacteria capable of memorizing transient sensory information for many cellular generations.

5.1 Investigation of the C-terminal tail of the cell cycle regulated DNA-(adenine N6)-methyltransferase from *Caulobacter crescentus*

The cell cycle regulated DNA-(adenine N6)-methyltransferase (CcrM) from *Caulobacter crescentus* is a well-studied enzyme in terms of biological function (Gonzalez et al. 2014), however, no structural information is available at the moment. CcrM is highly conserved in alpha-proteobacteria and essential in some species, for example in the plant pathogen *Agrobacterium tumefaciens* and the animal pathogen *Brucella abortus*. *Brucella abortus* can be transmitted to humans from sheep, cattle, or pigs (Robertson et al. 2000; Kahng and Shapiro 2001) and is a medical relevant pathogen, hence, making CcrM a potential drug target.

In the work presented here, the C-terminal domain of CcrM from *Caulobacter crescentus* was investigated. The C-terminal domain is conserved among CcrM homologs but its function was unknown so far. Deletion of the C-terminal domain lead to both, loss of catalytic activity and loss of DNA binding. The C-terminal domain alone is also not capable of DNA binding, but alanine mutations of conserved amino acids in the C-terminal domain of full length CcrM caused a drastic loss in catalytic activity with no changes in the secondary structures or cofactor binding behavior. However, dramatic changes in DNA binding of the mutant proteins could be observed, embodied either in increased DNA binding or for some mutant proteins in reduced DNA binding *in vitro*.

These results suggest that the C-terminal domain of *Caulobacter crescentus* CcrM is involved in DNA binding (Figure 15). The observed behavior of both, increased and decreased binding strength of different alanine mutants may be explained by different roles of the corresponding amino acid residues in the DNA binding process and DNA target sequence recognition. Amino acids that lead to a reduced DNA binding, when mutated to alanine, may be involved in direct or indirect interactions with DNA via hydrogen bonds. Amino acids that lead to an increased binding to DNA, when mutated to alanine, however, may be involved in the transmission of conformational changes of the CcrM-DNA complex upon target sequence recognition. These amino acid residues might be attributed to play a role in the so called indirect readout mechanism also described for the TATA box-binding protein (TBT), the trp repressor, the EcoRV endonuclease, or M.EcoRV (Otwinowski et al. 1988; Taylor et al. 1991; Bareket-Samish et al. 2000; Jurkowski et al. 2007). Sequence recognition by an indirect readout mechanism is based on the recognition of the DNA structure (bending, curvature, helical twist, flexibility) or the recognition of a change of the DNA conformation in the DNA-protein complex (Bareket-Samish et al. 2000; Zhang et al. 2004). In the case of CcrM, target sequence recognition could be explained by a two-step process. First, CcrM binds to DNA in a sequence independent manner and slides along the DNA. Then, upon target sequence recognition, a conformational change of the protein-DNA complex takes place, leading to an energetically preferred conformation. The ΔG release can then be used by the enzyme to undergo a conformational change, necessary for catalytic activity. If amino acid residues involved in this process are

mutated, this could lead to a very strong binding to DNA but adopting a conformation that is unsuitable for catalysis.

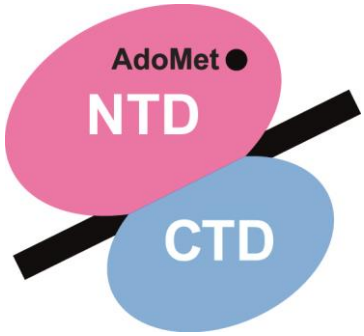


Figure 15. Hypothetical schematic view of CcrM bound to DNA (black line). The N-terminal domain (NTD) of CcrM is depicted in pink. The AdoMet binding site and the active site are located in the NTD. The C-terminal domain (CTD) is shown in blue. According to the experimental results described here, the CTD is involved in DNA binding as different mutations in this domain lead to both, increased and decreased binding behavior to DNA. (The image is taken from the manuscript attached as Appendix 1.)

Bacterial DNA C-MTases and N-MTases share a common architecture comprising a catalytic domain, which includes the active site and AdoMet-binding site and a target recognition domain (TRD), involved in the recognition of the specific DNA target sites (Malone et al. 1995; Jeltsch 2002). The catalytic region and the AdoMet binding site are composed of up to ten conserved motifs (I to X). The occurrence of different orders of the motifs in different MTases are a result of circular permutations that emerged during molecular evolution. Besides the different orders of the ten conserved motifs the location of the target recognition region also differs (Figure 4). Three different arrangements have been found so far for adenine N6 MTases (α, β, γ), however, theoretically, there are three more arrangements possible (δ, ε, ζ).

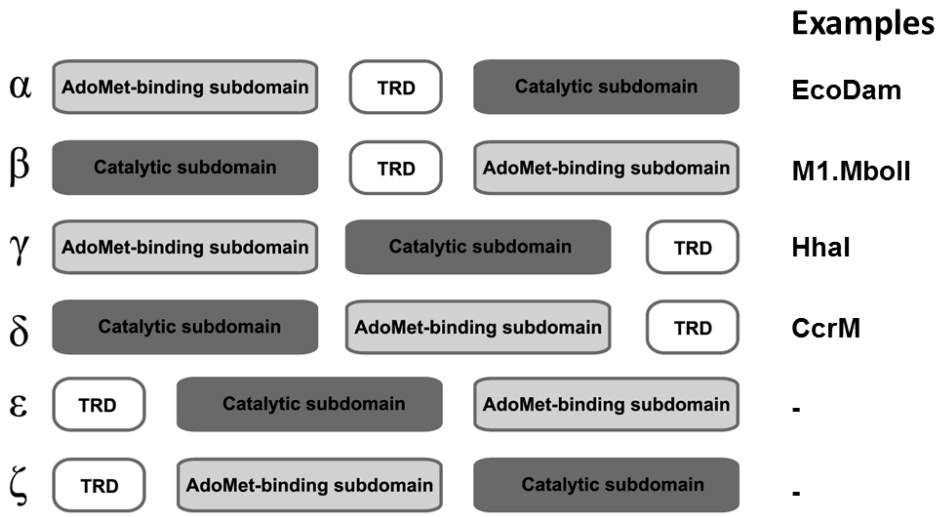


Figure 16. Different arrangements of the three conserved regions of bacterial DNA MTases with examples of α, β, and γ DNA MTases. Re-classification of CcrM from the β to the δ-class has been proposed based on the results presented in this thesis. (The image is modified from the manuscript attached as Appendix 1. The figure legend has been modified from Appendix 1.)

The involvement of the C-terminal domain of CcrM in DNA binding and target sequence recognition suggests that the C-terminal domain of CcrM is the target recognition domain. Following the classification of methyltransferases by Malone et al. from 1995, CcrM would be the first example of the δ class of methyltransferases (Figure 4) (Malone et al. 1995). It will be interesting to see, if this conclusion drawn from the discussed biochemical studies can be verified by structural data. Therefore, e.g. a protein crystal structure of *Caulobacter crescentus* CcrM or one of its homologs in complex with DNA is eagerly awaited.

5.2 Design of synthetic epigenetic circuits in bacteria

The field of synthetic biology is striving for genetically programming cells or organisms to perform a large variety of desired tasks. In this part of the project, the bacterial DNA-(adenine N6)-methyltransferase CcrM from *Caulobacter crescentus* and a designed zinc finger protein have been used to create a synthetic epigenetic memory system.

5.2.1 Building synthetic systems featuring DNA methylation

First, there was the idea to build synthetic biology systems harnessing DNA methylation as a control signal to regulate gene expression in *Escherichia coli*. The used CcrM methyltransferase methylates adenine residues in 5'-GANTC-3' sequences. In *Escherichia coli*, there are no methyltransferases known that methylate this target site. Hence, none or only minor burdening of the host was expected upon CcrM expression, what was also confirmed experimentally (Supplementary Figure 1A and Supplementary Figure 16 from Appendix 2). In contrast to eukaryotic systems, in which DNA-(cytosine C5)-methylation of promoter regions is a repressive epigenetic mark (Suzuki and Bird 2008), in the herein developed system DNA methylation acts as an activating mark, because the methylation of promoter regions is intended to interfere with transcriptional repressor binding and leads to gene expression. Zinc finger proteins were used as transcriptional repressors due to their programmability to bind to freely selectable target sequences and their sensitivity of DNA binding to DNA methylation. Zinc finger domains recognize AT base pairs via asparagine or glutamine residues and make specific contacts to the adenine base via the N6 and N7 position (Seeman et al. 1976; Aggarwal et al. 1988) (Supplementary Figure 2A from Appendix 2). Hence, adenine-N6 methylation should impair the recognition of AT base pairs. Multiple C2H2 zinc finger proteins consisting of three zinc finger domains with target sequences that overlapped with the CcrM target sequence (5'-GANTC-3') were designed and DNA binding was tested with unmethylated and methylated binding sites. Reduced DNA binding was observed upon DNA methylation for all tested proteins and the most promising zinc finger protein candidate was used for further experimental procedures. A synthetic promoter repressor system was constructed and incrementally extended into the final epigenetic memory system.

As a first system, a GFP reporter gene was set under the control of a constitutive promoter. Then, zinc finger binding sites were inserted upstream and downstream of the -35 region of the promoter and upon expression of a dimerized version of the zinc finger protein, gene repression could be observed. Upon expression of the CcrM methyltransferase, and subsequent methylation of the zinc finger binding sites, GFP expression could be restored.

The next step was to regulate the expression of CcrM itself by the methylation dependent gene regulation system and thereby implementing a positive feedback system. To this end, the gene for CcrM was cloned downstream of the GFP reporter gene with an additional ribosomal binding site, resulting in an artificial two-gene operon. Theoretically, this system should feature bistability. The system should be repressed by the zinc finger protein (off-state) and once the operon gets expressed the system should stay in an on-state, due to constant methylation of the promoter region. In order to achieve this desired behavior, it was necessary to introduce additional zinc finger binding sites to maintain a stable off-state, as spontaneous CcrM expression would switch the system to the on-state unintendedly due to the positive feedback. Additionally, regulation of the expression of the zinc finger repressor by methylation-independent autoregulation was necessary to reduce fluctuations of the repressor protein levels that would lead to cessation of the positive feedback signal in incidences of very high repressor levels.

The intrinsic problem of spontaneous on switching of systems with positive feedback was approached from two sides. Very tight repression of the system was achieved by introduction of additional repressor binding sites, as described above. Moreover, the reversibility of epigenetic systems was exploited and a reset-switch was developed. The reset-switch works by an orthogonal protein degradation system, allowing for inducible degradation of CcrM. This leads to passive loss of promoter methylation and hence leads to a stable off-state.

According to literature database search, the developed system described here, represents the first report on a synthetic epigenetic memory system based on DNA-(adenine N6)-methylation. The system can be used in different setups for the detection and subsequent memorization of different physical and chemical signals (heat, arabinose, cisplatin, UV radiation).

5.2.2 Potential biotechnological applications of the synthetic epigenetic memory system

The developed system demonstrates that epigenetic mechanisms can be applied in bacterial synthetic biology approaches. Systems using the epigenetic memory system could find applications in various fields of biotechnology and a selection of examples is described below.

The described epigenetic memory system is composed of a memory module and a trigger module. This modularity offers the opportunity to couple many more trigger signals to the memory module. Examples of existing and described sensory parts that could be easily coupled to the system are sensors that react on antibiotics like tetracycline (TetR) (Korpela et al. 1998) or cellular signals found in communicating bacteria by quorum sensing (LuxR) (Wu et al. 2000). Additionally, ribozymes could be used for example as fluoride sensors (Baker et al. 2012). Furthermore, there have been reports on designed ribozyme functions, e.g. shown for TNT detection (Davidson et al. 2012).

In the current set up of the epigenetic memory system, one single signal can be processed and stored at a time. In order to extend the system so that several different input signals can be processed and stored at the same time, additional DNA MTases could be used simultaneously that store information in distinguishable DNA methylation patterns. Additional to CcrM, which is used in the current system, M.HindIII and M.XmnI could be used even by employing the same ZnF repressor protein. Protein sequences of both methyltransferases are available (GenBank: CBW29751.1 and GenBank: AAC44403.1) and genes coding for M.HindIII and M.XmnI could be easily synthesized. In the current memory system, the used ZnF protein binds to 5'-GGAGAAG**A**A-3' sequences (the bold adenine base can be methylated and thereby modulates the ZnF binding). The ZnF binding site overlaps with the target sequence of CcrM (5'-G**A**NTC-3') resulting in 5'-GGAGAAG**A**ATC-3' (the ZnF binding site is underlined, the CcrM target site is overlined). In order to use M.HindIII (5'-**A**AGCTT-3'), one could create ZnF binding sites that overlap with the MTase target site, resulting in 5'-GGAGAAG**A**AGCTT-3'. For the MTase M.XmnI (5'-G**A**ANNNNTTC-3'), the combined DNA sequence would be 5'-GGAGAAG**A**ANNNNTTC-3'. If successful, this would allow for storing simultaneously information of three different input signals. In

order to use other DNA MTases for signal storage, new DNA methylation sensitive repressor modules could be engineered based on zinc finger proteins or TALEs (Bultmann et al. 2012; Deng et al. 2012). Examples of potential methyltransferases that are likely to work well in *E. coli* are M.EcoRI and M.EcoRV (Reich and Mashhoon 1991; Jeltsch et al. 1998). These two enzymes are naturally present in certain *E. coli* strains, hence no adverse effects on the host cells are expected upon using these enzymes.

Once at least two compatible memory systems are established, multi-input systems could be created, that would be of value in the field of biological computing and in systems that would integrate input signals based on Boolean logic could be created (Figure 17).

Initial data suggests that the epigenetic memory system sensitive to arabinose might work already as an OR gate (data not shown). In this setup, the trigger module reacts on arabinose and thereby can switch the system to the on-state. Additionally, the memory module has the intrinsic property to be switched to the on-state upon heat induction due to thermal sensitivity of the zinc finger repressor. Thus, arabinose or heat, or both signal can switch the system to the on-state.

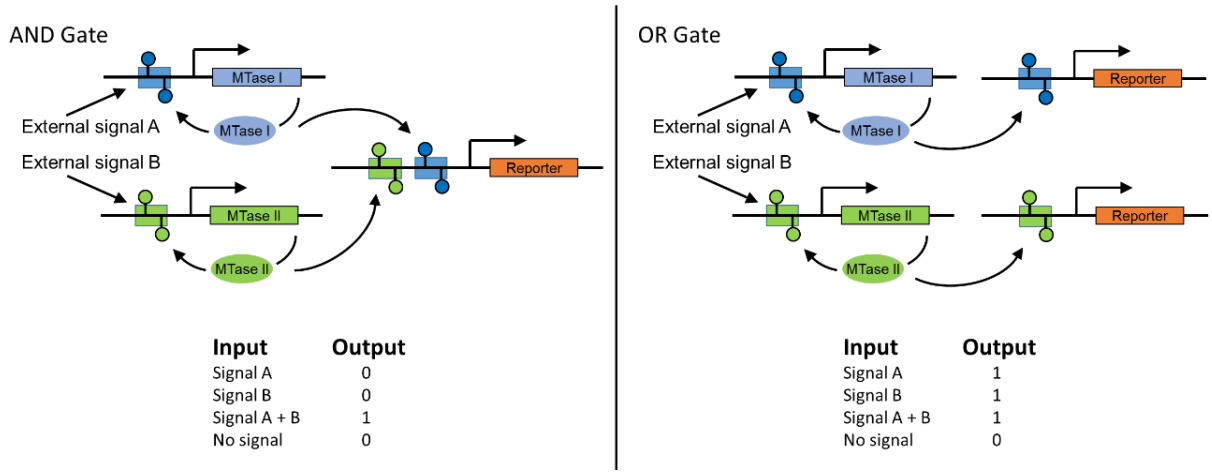


Figure 17. Exemplary epigenetic circuits featuring logic gates that integrate two input signals. For the depicted AND gate, signals A and B have to be present in order to create an output signal (expression of the reporter gene). The illustrated OR gate gives an output signal when either signal A, B or both signals are present.

Sequencing based readout systems that are able to read DNA methylation patterns, independent of reporter genes, would be advantageous especially when using more than one methyltransferase in the system. New sequencing techniques based on nanopore sequencing or single-molecule real time (SMRT) sequencing could be used and thereby would release the restriction to a limited number of reporter proteins (Flusberg et al. 2010; Davis et al. 2013; Rand et al. 2017; Simpson et al. 2017). Complex methylation patterns could be easily read even at the level of single cells due to individual read lengths in the range of several kilo bases.

Other synthetic biology approaches for storing external information in living bacteria described in the literature are mainly built on gene regulation by bacterial transcription factors or recombinase based systems. The first of which, ignited the field of synthetic biology and uses two transcriptional repressor proteins to establish a toggle switch (Gardner et al. 2000) (Figure 1A). In systems developed using recombinases, pieces of DNA are inverted upon a trigger signal and thereby generate an output signal (Siuti et al. 2013). Memory created by recombinases does not need active maintenance, however, using recombinases might have adverse effects on host genome integrity. There are natural model systems for both mechanisms, e.g. the phage lambda system, with lambda repressor and Cro-protein that negatively affect each other and phase variation in *Escherichia coli* by *fim* promoter inversion (Klemm 1986; Ptashne 2004). Besides the use of additional DNA MTases for storing information in live bacteria, other approaches like recombinase based systems may be used as complementary information storage units in order to increase complexity of the circuits (Yang et al. 2014). In fine-tuned systems, that use a multitude of information storage mechanisms, a large number of signals could be processed at the same time and complex logical operations could be conducted.

The current epigenetic memory system uses DNA methylation as an activating mark for bacterial gene expression via interference with transcriptional repressor binding. However, it would be also conceivable to construct synthetic epigenetic systems that use DNA methylation as a repressive mark for the regulation of gene expression in bacteria. Systems that resemble the described bacterial two-hybrid system used here for zinc finger design (Wright et al. 2006) could be constructed. Here, a zinc finger fusion protein to Gal11P and a RNA polymerase fusion to Gal4 could be applied in a

way that zinc finger binding in a crippled promoter region would recruit RNA polymerase and initiate transcription (Figure 18A). Zinc finger binding would be disturbed by methylation of the zinc finger binding site and thereby transcription would stop. Using a system in which a methyltransferase regulates its own expression by promoter methylation, an oscillating expression profile is expected, the frequency of which would reflect the cell division rate (Figure 17B). Using three MTases that negatively regulate each other, construction of oscillating epigenetic circuits would be conceivable, similar to the repressilator described by Elowitz et al. (Elowitz and Leibler 2000) (Figure 18C and D).

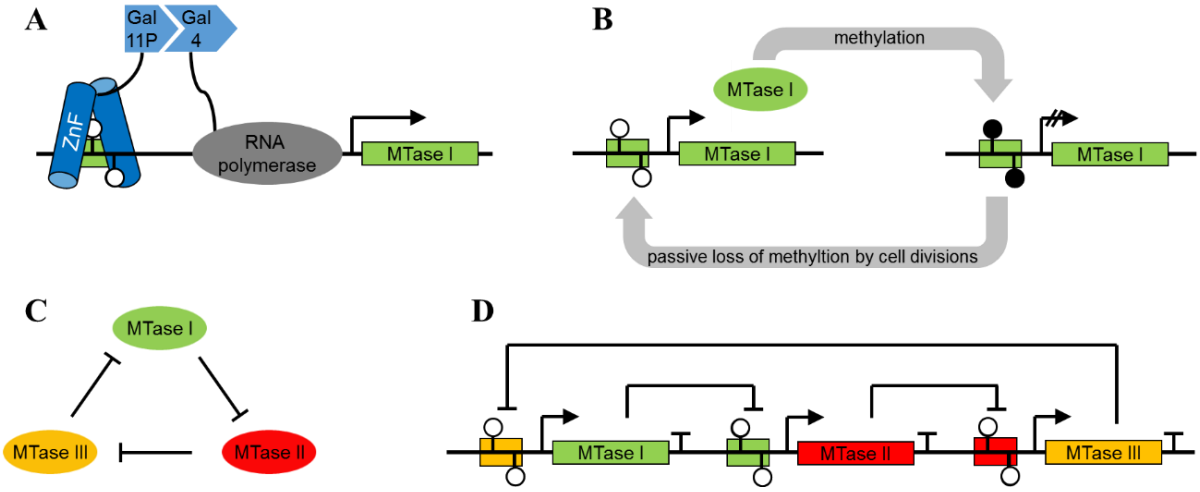


Figure 18. Potential circuit design in which DNA methylation leads to gene repression and would allow for the construction of oscillating epigenetic circuits. A) Methylation sensitive zinc finger protein fused to Gal11P protein, which can interact with Gal4, fused to an RNA polymerase molecule. DNA binding of the zinc finger protein leads to recruitment of RNA polymerase and subsequently to gene activation. No gene activation would occur upon methylation of the zinc finger binding site. The design idea is adapted from a bacterial two-hybrid system (Wright et al. 2006). B) Potential self-regulating circuit that would oscillate following cell divisions. The MTase would be expressed if the zinc finger binding site is unmethylated. Then, the MTase methylates the zinc finger binding site and transcription stops. Passive loss of methylation by DNA replication then leads again to MTase expression. C) By constructing a circuit that would consist of three MTases that negatively regulate each other, another epigenetic oscillating system could be constructed. D) Different representation of the oscillating epigenetic circuit using three MTases. MTaseI would negatively regulate MTaseII, MTaseII would negatively regulate MTaseIII, MTaseIII would negatively regulate MTaseI. Upon initial induction of one of the MTases, an oscillating expression pattern of the three MTases is expected.

One near term application goal could be the development of a biological sensor based on the existing heat inducible system that can be used for monitoring of cold chains.

With the current settings, temperature changes from 30°C to 37°C induce the memory system and this information can be stored in the bacterial memory system for many cellular generations. To broaden the applicability of the temperature senor, and to create sensors that switch at other temperature thresholds, mutant libraries of the zinc finger protein could be screened for desired behavior. For industrial applications, bacterial cultures could be encapsulated in buffered culture medium and would be independent devices without the need for electricity or electronic components. Additionally, coupling of the biosensor to a readout system, similar to immunological pregnancy tests, would ease the use and would make the readout very simple. The feature of reversibility of the epigenetic memory system could be used in order to ensure off-state conditions at the beginning of a monitoring period.

An additional potential biotechnological application would be the use of the epigenetic system with positive feedback as a biocontainment system. A biocontainment system could be achieved by repressing an operon coding for a maintenance methyltransferase and a gene for a toxic product. Toxic gene products could be for example restriction endonucleases, lysozyme, or highly reactive proteases. Under laboratory conditions, the system would be kept in the off-state, once the bacteria harboring this containment system would leave the laboratory conditions unintendedly, the system would switch to the on-state and the methyltransferase and the toxic genes would be expressed, leading to cell death. Systems described in literature use for example synthetic protein design based on non-standard amino acids and demand for extensive editing of the *E. coli* genome but being also very effective (Mandell et al. 2015; Rovner et al. 2015). Synthetic monostable circuits have been also designed to work as a biocontainment system (Chan et al. 2016). However, systems expressing toxic gene that kill bacteria upon a certain signal poses evolutionary pressure on the bacteria. Therefore, in the setup described by Chan et al., it was necessary to use several toxic products in combination to efficiently kill bacterial cells on purpose. Possibly this would also apply for a biocontainment system based on the epigenetic memory system described here and would demand for a system based on multiple toxic gene products acting in concert.

Another application of the epigenetic memory system could be the use of the system as a protein expression platform. In such a setting, the epigenetic memory system

would obviate the need for a constant inductor supplementation of the bacterial induction strain. Only a transient signal would be needed to induce the system and subsequently the protein production would be perpetuated. In this setup, the maintenance operon would consist of a methyltransferase and the gene for the protein to be expressed.

The existing CcrM based epigenetic memory system has great potential to be used in a range of industrial applications. However, a more detailed characterization of the system and the switching behavior is needed before such applied systems can be developed. This includes analysis of spontaneous on-switching, behavior of the system under a broader set of environmental conditions and the investigation of long-term stability of on- and off-states. The thorough characterization of the systems allows then also for comparison of the epigenetic system with other existing synthetic memory systems popular in the field of synthetic biology for example based on recombinases or transcription factors (Kobayashi et al. 2004; Friedland et al. 2009).

5.2.3 Modelling of the epigenetic memory system

Examining the developed epigenetic memory system with a systems biology approach would aid to shed light on issues like stability of the two states and potential weak points of the systems that could be adjusted and improved. For computational modelling, certain system parameters have to be experimentally determined that are indicated in Figure 19. This includes the protein concentrations, synthesis rates and degradation rates of CcrM, EGFP and the zinc finger repressor protein. Protein concentrations can be easily determined for example by western blotting (Charette et al. 2010). Synthesis and degradation rates can be assessed by ribosome profiling (Li et al. 2014) and pulse-chase analysis (Simon and Kornitzer 2014). Additionally, the methylation activity of CcrM *in vivo* has to be determined. This could be derived from the global methylation state at 5'-GANTC-3' sites, at a known CcrM concentration and the known cell division rate. The global methylation state could be assessed by a methyl acceptance assay (Nephew et al. 2009). Absolute DNA binding strength of the zinc finger repressor protein to methylated and unmethylated binding sites have to be quantified. In the scope of this thesis, a roughly 70 fold preference for unmethylated DNA has been measured (Supplementary Figure 5 from Appendix 2). With the present

zinc finger protein purification settings, it was not possible to determine absolute DNA binding constants, due to DNA co-purification. Protein purification has to be optimized in order to obtain pure zinc finger protein without DNA to determine the binding constants.

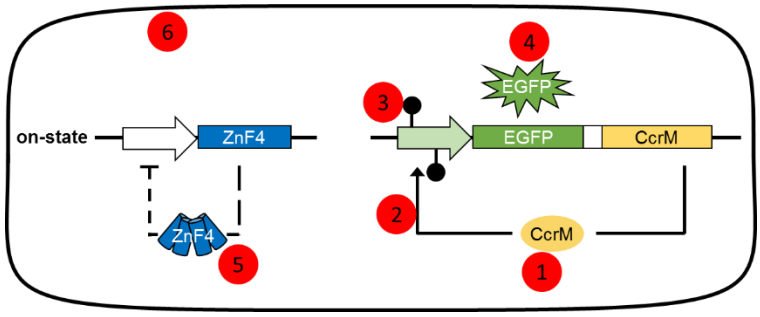


Figure 19. Scheme of the synthetic epigenetic memory system in the on-state. Indicated with numbered red circles are system properties that need to be determined for mathematical modelling of the system. 1: CcrM protein concentration in the on- and off-state. 2: DNA methylation rate of CcrM. 3: methylation status of the promoter region. 4: EGFP protein concentration in on- and off-states. 5: ZnF protein concentration. 6: cell division rate. Additionally, protein synthesis rates as well as protein degradation rates for CcrM and EGFP have to be determined.

Based on the following differential equations, a basic mathematical model of the epigenetic memory system could be compiled. The expression of the zinc finger protein with autoregulation could be described by two equations, which describe the change of the zinc finger protein concentration over time (1) and the change of zinc finger mRNA concentration over time (2).

$$\frac{dp_{ZnF}}{dt} = L\, m_{ZnF} - D_{p_{ZnF}}\, p_{ZnF} - \frac{1}{2}\mu\, p_{ZnF} \qquad (1)$$

$$\frac{dm_{ZnF}}{dt} = T\,(p_{ZnF}) - D_{m_{ZnF}}\, m_{ZnF} - \frac{1}{2}\mu\, m_{ZnF} \qquad (2)$$

where p_{ZnF} is the protein concentration of the zinc finger repressor, L is the translation rate, m_{ZnF} is the mRNA concentration of the zinc finger repressor transcript, $D_{p_{ZnF}}$ is the protein degradation rate of the zinc finger repressor protein, μ is the growth rate, T is the transcription rate of the zinc finger repressor gene, $D_{m_{ZnF}}$ is the degradation rate of the zinc finger repressor mRNA transcript.

The change in zinc finger protein concentration ($\frac{dp_{ZnF}}{dt}$) equals the translation ($L\, m_{ZnF}$), minus the zinc finger protein degradation ($D_{p_{ZnF}}\, p_{ZnF}$) and the dilution of the zinc finger

protein concentration by cell division ($\frac{1}{2}\mu\, p_{ZnF}$) (Equation 1). The change in zinc finger mRNA concentration equals the transcription ($T\ (p_{ZnF})$), which is negatively influenced by the zinc finger protein concentration, minus the zinc finger mRNA degradation ($D_{m_{ZnF}}\, m_{ZnF}$) and the dilution of zinc finger mRNA concentration by cell division ($\frac{1}{2}\mu\, m_{ZnF}$) (Equation 2). Due to the negative influence of the zinc finger protein on zinc finger transcription, a constant concentration should be attuned.

A simplified description of the changes in proteins encoded by the reporter-maintenance operon and the methylation state of the promoter of the reporter-maintenance operon is given by equations (3), (4), and (5).

$$\frac{dp_{CcrM}}{dt} = TL(\theta) - D_{p_{CcrM}}\, p_{CcrM} - \frac{1}{2}\mu\, p_{CcrM} \qquad (3)$$

$$\frac{d\theta}{dt} = M\, p_{CcrM}(1-\theta) - \frac{1}{2}\mu\, \theta \qquad (4)$$

$$\frac{dp_{EGFP}}{dt} = TL(\theta) - D_{p_{EGFP}}\, p_{EGFP} - \frac{1}{2}\mu\, p_{EGFP} \qquad (5)$$

where p_{CcrM} is the protein concentration of CcrM, $TL(\theta)$ is the protein production as a function of time dependent on the methylation state status (θ), $D_{p_{CcrM}}$ is the degradation rate of CcrM, μ is the growth rate, M is the methylation rate of CcrM, p_{EGFP} is the protein concentration of EGFP, and $D_{p_{EGFP}}$ is the degradation rate of EGFP.

The change of CcrM protein concentration ($\frac{dp_{CcrM}}{dt}$) would equal the protein production ($TL(\theta)$), which is positively influenced by the methylation status (θ), minus CcrM degradation ($D_{p_{CcrM}}\, p_{CcrM}$) and the dilution of CcrM by cell division ($\frac{1}{2}\mu\, p_{CcrM}$) (Equation 3). The change in the methylation state ($\frac{d\theta}{dt}$) equals the methylation rate of CcrM ($M\, p_{CcrM}(1-\theta)$), minus the passive loss of methylation by DNA replication, simplified to be dependend on cell division ($\frac{1}{2}\mu\, \theta$) (Equation 4). The change in EGFP protein concentration over time ($\frac{dp_{EGFP}}{dt}$) equals the formation dependent on the methylation status ($TL(\theta)$), minus the EGFP degradation ($D_{p_{EGFP}}\, p_{EGFP}$) and the dilution of EGFP concentration by cell division ($\frac{1}{2}\mu\, p_{EGFP}$) (Equation 5).

Mathematical modelling of the epigenetic memory system would lead to a better understanding of the system and would aid the construction of more and more complex circuits as demonstrated for automated genetic circuit design (Ellis et al. 2009; Nielsen et al. 2016).

5.3 Identification, cloning and heterologous expression of *Citrobacter* sp. SG hydrogenase 2

Molecular hydrogen is an attractive energy carrier, due to a high energy per mass content and combustion products being only water and heat under ideal conditions. However, the main route of molecular hydrogen production today is steam reforming of natural gas (Abbas and Daud 2010). In order to use hydrogen as a true environmentally friendly energy carrier, the need for fossil fuels for hydrogen production has to be omitted. Several attempts have been made into this direction. Examples are electrolysis of water by using electricity generated from “renewable energy” (Wang et al. 2014), polymeric catalysts that produce hydrogen by water splitting using sunlight as energy source (Wang et al. 2009b) and progress is made in the field of biohydrogen production (Hallenbeck et al. 2012). Using a synthetic biology approach, that would couple hydrogen production to photosynthesis and thereby use sunlight as primary energy source would be a scalable and attractive approach. This approach, however, is limited mainly by the lack of appropriate enzymes capable of hydrogen production in the presence of molecular oxygen, which inevitably emerges by photosynthesis.

In the scope of this thesis, a *Citrobacter* species, which produces considerable amounts of molecular hydrogen at ambient culturing conditions was discovered. This is of great interest, as most known hydrogen producing enzymes are inactivated or destroyed in the presence of molecular oxygen (Vignais and Billoud 2007).

As the main hydrogen detecting tool, an unconventional and easily operatable hydrogen detector was used. The detector is a semiconducting device, usually applied in mechanical engineering to detect very small leaks in tubing systems by flushing them with a hydrogen containing gas mixture. This device was used as an easy to handle, low-cost, quantitative and reliable hydrogen detector. The system was

validated by gas chromatography and could be an interesting tool for biotechnological research in the hydrogen research field.

The hydrogen producing *Citrobacter* species was fortuitously discovered and was named *Citrobacter* sp. SG (Stuttgart, Germany). Next, the enzyme responsible for hydrogen production was identified by applying a set of different protein purification steps followed by peptide mass fingerprinting. The enzyme was determined to be a type 2 [NiFe] hydrogenase. The hydrogenase 2 operon was cloned into an inducible expression vector and it was possible to express the active enzyme in *E. coli* cells in a nickel dependent manner.

Interestingly, a Japanese group of researchers reported similar findings on an oxygen tolerant hydrogenase from another *Citrobacter* sp. (*Citrobacter* sp. S-77) (Eguchi et al. 2012). Later, the group revealed that the hydrogen producing enzyme is a type 2 hydrogenase and thereby support the results presented here (Muhd Noor et al. 2016). Sequence comparison revealed highly similar amino acid composition of hydrogenase 2 enzymes of *Citrobacter* sp. SG, *Citrobacter* sp. S-77 and *Citrobacter freundii*, leading to the hypothesis that the oxygen tolerance is not only limited to *Citrobacter* sp. SG or *Citrobacter* sp. S-77 but rather a C*itrobacter* specific characteristic.

Bioinformatic sequence comparison of small and large subunits of the *Citrobacter* sp. SG enzyme with the known oxygen tolerant hydrogenase 1 from *E. coli* revealed that a novel mechanism has to be responsible for oxygen tolerance of *Citrobacter* sp. SG hydrogenase 2. *E. coli* hydrogenase 1 and other known oxygen tolerant membrane bound hydrogenases comprise an unusual [4Fe3S] cluster that is coordinated by six cysteine residues in the small subunit, located proximal to the active site in the large subunit. The [4Fe3S] cluster is responsible for the donation of electrons in case of inactivation by oxygen (Fritsch et al. 2011; Frielingsdorf et al. 2014)(see chapter 1.3.4).

In the *Citrobacter* sp. hydrogenase 2 small subunits, no additional cysteine residues can be found that could be attributed to play a role in coordination of the unusual [4Fe3S] cluster proximal to the active site. However, there are two additional cysteine residues in the large subunits of the *Citrobacter* sp. (Appendix 5). These cysteine residues are most probably not involved in the same mechanism observed in *E. coli* hydrogenase 1 like enzymes. Mapping of the residues to the *E. coli* hydrogenase 1

crystal structure suggests that they are located far away from the small subunit (Figure 14). Potential involvement of the additional cysteine residues of *Citrobacter* sp. hydrogenase 2 large subunit in the formation of quaternary structure is conceivable. Such larger structures may aid to form a microenvironment, in which low oxygen levels can be maintained and this may help to keep the hydrogenases active in presence of oxygen. Alternatively, intermolecular complexes with other proteins could be formed that would help to tolerate oxygen present in the environment, similarly as also described for hydrogenase 1 from *E. coli* (Wulff et al. 2016).

The discovered *Citrobacter* sp. SG produces hydrogen at ambient conditions. This suggests that the regulation of hydrogenase 2 in *Citrobacter* sp. might be different from the regulation in *E. coli*, where hydrogenase 2 is only expressed under anaerobic conditions. In *E. coli*, hydrogenase 2 is used for H_2 uptake and H_2 in turn is used as an electron source in anaerobic respiration. In this process, electrons are transferred via the quinone pool to fumarate as the final electron acceptor (Kröger et al. 1992; Sawers 1994; Richard et al. 1999). In contrast, in *Citrobacter* species, hydrogenase 2 could be involved in adaptations to acidic growth conditions and help to maintain less acidic intracellular conditions by reducing protons to molecular hydrogen. This process would be very efficient, as gaseous H_2 is virtually withdrawn from chemical equilibrium.

Potentially, the discovered enzyme could be applied biotechnologically in the field of hydrogen production by coupling of photosynthesis and H_2 production. This together with the constant progress in technologies related to the storage of hydrogen could promote the usability and prevalence of hydrogen as the future energy carrier (Hu et al. 2016; Sethia and Sayari 2016; Kothandaraman et al. 2017).

6 References

Abbas HF, Daud WW. 2010. Hydrogen production by methane decomposition: a review. *international journal of hydrogen energy* **35**: 1160-1190.

Aggarwal AK, Rodgers DW, Drottar M, Ptashne M, Harrison SC. 1988. Recognition of a DNA operator by the repressor of phage 434: a view at high resolution. *Science* **242**: 899.

Albu RF, Jurkowski TP, Jeltsch A. 2012a. The Caulobacter crescentus DNA-(adenine-N6)-methyltransferase CcrM methylates DNA in a distributive manner. *Nucleic acids research* **40**: 1708-1716.

Albu RF, Zacharias M, Jurkowski TP, Jeltsch A. 2012b. DNA interaction of the CcrM DNA methyltransferase: a mutational and modeling study. *Chembiochem* **13**: 1304-1311.

Allis CD, Jenuwein T. 2016. The molecular hallmarks of epigenetic control. *Nature Reviews Genetics* **17**: 487-500.

Annaluru N, Muller H, Mitchell LA, Ramalingam S, Stracquadanio G, Richardson SM, Dymond JS, Kuang Z, Scheifele LZ, Cooper EM. 2014. Total synthesis of a functional designer eukaryotic chromosome. *Science* **344**: 55-58.

Antal T, Krendeleva T, Laurinavichene T, Makarova V, Ghirardi M, Rubin A, Tsygankov A, Seibert M. 2003. The dependence of algal H 2 production on photosystem II and O 2 consumption activities in sulfur-deprived Chlamydomonas reinhardtii cells. *Biochimica et Biophysica Acta (BBA)-Bioenergetics* **1607**: 153-160.

Arber W, Dussoix D. 1962. Host specificity of DNA produced by Escherichia coli: I. Host controlled modification of bacteriophage λ. *Journal of molecular biology* **5**: 18-36.

Armaroli N, Balzani V. 2011. The hydrogen issue. *ChemSusChem* **4**: 21-36.

Baker JL, Sudarsan N, Weinberg Z, Roth A, Stockbridge RB, Breaker RR. 2012. Widespread genetic switches and toxicity resistance proteins for fluoride. *Science* **335**: 233-235.

Bareket-Samish A, Cohen I, Haran TE. 2000. Signals for TBP/TATA box recognition. *Journal of molecular biology* **299**: 965-977.

Barthelemy H, Weber M, Barbier F. 2016. Hydrogen storage: Recent improvements and industrial perspectives. *international journal of hydrogen energy*.

Basak N, Das D. 2007. The prospect of purple non-sulfur (PNS) photosynthetic bacteria for hydrogen production: the present state of the art. *World Journal of Microbiology and Biotechnology* **23**: 31-42.

Basu S, Gerchman Y, Collins CH, Arnold FH, Weiss R. 2005. A synthetic multicellular system for programmed pattern formation. *Nature* **434**: 1130-1134.

Baylin SB, Jones PA. 2011. A decade of exploring the cancer epigenome—biological and translational implications. *Nature Reviews Cancer* **11**: 726-734.

Berens C, Groher F, Suess B. 2015. RNA aptamers as genetic control devices: the potential of riboswitches as synthetic elements for regulating gene expression. *Biotechnology journal* **10**: 246-257.

Bergman Y, Cedar H. 2013. DNA methylation dynamics in health and disease. *Nature Structural & Molecular Biology* **20**: 274-281.

Bhaya D, Davison M, Barrangou R. 2011. CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annual review of genetics* **45**: 273-297.

Bickle TA, Krüger D. 1993. Biology of DNA restriction. *Microbiological reviews* **57**: 434-450.

Bikard D, Jiang W, Samai P, Hochschild A, Zhang F, Marraffini LA. 2013. Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic acids research* **41**: 7429-7437.

Blokesch M, Paschos A, Theodoratou E, Bauer A, Hube M, Huth S, Böck A. 2002. Metal insertion into NiFe-hydrogenases. *Biochemical Society Transactions* **30**: 674-680.

Böck A, King PW, Blokesch M, Posewitz MC. 2006. Maturation of hydrogenases. *Advances in microbial physiology* **51**: 1-225.

Boyd E, Peters JW. 2013. New insights into the evolutionary history of biological nitrogen fixation. *Frontiers in microbiology* **4**: 201.

Brophy JA, Voigt CA. 2014. Principles of genetic circuit design. *Nature methods* **11**: 508-520.

Buhrke T, Lenz O, Krauss N, Friedrich B. 2005. Oxygen tolerance of the H2-sensing [NiFe] hydrogenase from Ralstonia eutropha H16 is based on limited access of oxygen to the active site. *Journal of Biological Chemistry* **280**: 23791-23796.

Bultmann S, Morbitzer R, Schmidt CS, Thanisch K, Spada F, Elsaesser J, Lahaye T, Leonhardt H. 2012. Targeted transcriptional activation of silent oct4 pluripotency gene by combining designer TALEs and inhibition of epigenetic modifiers. *Nucleic acids research* **40**: gks199.

Chan CT, Lee JW, Cameron DE, Bashor CJ, Collins JJ. 2016. 'Deadman'and'Passcode'microbial kill switches for bacterial containment. *Nature chemical biology* **12**: 82-86.

Charette SJ, Lambert H, Nadeau PJ, Landry J. 2010. Protein quantification by chemiluminescent Western blotting: elimination of the antibody factor by dilution series and calibration curve. *Journal of immunological methods* **353**: 148-150.

Cheng L, Keiler KC. 2009. Correct timing of dnaA transcription and initiation of DNA replication requires trans translation. *Journal of bacteriology* **191**: 4268-4275.

Cheng X. 1995. DNA modification by methyltransferases. *Current opinion in structural biology* **5**: 4-10.

Cheng X, Kumar S, Posfai J, Pflugrath JW, Roberts RJ. 1993. Crystal structure of the Hhal DNA methyltransferase complexed with S-adenosyl-L-methionine. *Cell* **74**: 299-307.

Cheng X, Roberts RJ. 2001. AdoMet-dependent methylation, DNA methyltransferases and base flipping. *Nucleic acids research* **29**: 3784-3795.

Colbeau A, Kelley BC, Vignais PM. 1980. Hydrogenase activity in Rhodopseudomonas capsulata: relationship with nitrogenase activity. *Journal of bacteriology* **144**: 141-148.

Collier J. 2012. Regulation of chromosomal replication in Caulobacter crescentus. *Plasmid* **67**: 76-87.

Collier J, McAdams HH, Shapiro L. 2007. A DNA methylation ratchet governs progression through a bacterial cell cycle. *Proceedings of the National Academy of Sciences* **104**: 17111-17116.

Collier J, Murray SR, Shapiro L. 2006. DnaA couples DNA replication and the expression of two cell cycle master regulators. *The EMBO journal* **25**: 346-356.

Collier J, Shapiro L. 2009. Feedback control of DnaA-mediated replication initiation by replisome-associated HdaA protein in Caulobacter. *Journal of bacteriology* **191**: 5706-5716.

Curtis PD, Brun YV. 2010. Getting in the loop: regulation of development in Caulobacter crescentus. *Microbiology and Molecular Biology Reviews* **74**: 13-41.

Danino T, Mondragón-Palomino O, Tsimring L, Hasty J. 2010. A synchronized quorum of genetic clocks. *Nature* **463**: 326-330.

Davidson ME, Harbaugh SV, Chushak YG, Stone MO, Kelley-Loughnane N. 2012. Development of a 2, 4-dinitrotoluene-responsive synthetic riboswitch in E. coli cells. *ACS chemical biology* **8**: 234-241.

Davis BM, Chao MC, Waldor MK. 2013. Entering the era of bacterial epigenomics with single molecule real time DNA sequencing. *Current opinion in microbiology* **16**: 192-198.

Deng D, Yin P, Yan C, Pan X, Gong X, Qi S, Xie T, Mahfouz M, Zhu J-K, Yan N. 2012. Recognition of methylated DNA by TAL effectors. *Cell research* **22**: 1502.

Din MO, Danino T, Prindle A, Skalak M, Selimkhanov J, Allen K, Julio E, Atolia E, Tsimring LS, Bhatia SN. 2016. Synchronized cycles of bacterial lysis for in vivo delivery. *Nature* **536**: 81-85.

Dincer I, Acar C. 2015. Review and evaluation of hydrogen production methods for better sustainability. *international journal of hydrogen energy* **40**: 11094-11111.

Domian IJ, Reisenauer A, Shapiro L. 1999. Feedback control of a master bacterial cell-cycle regulator. *Proceedings of the National Academy of Sciences* **96**: 6648-6653.

Dominguez AA, Lim WA, Qi LS. 2015. Beyond editing: repurposing CRISPR-Cas9 for precision genome regulation and interrogation. *Nature reviews Molecular cell biology* **17**: 5-15.

Duché O, Elsen S, Cournac L, Colbeau A. 2005. Enlarging the gas access channel to the active site renders the regulatory hydrogenase HupUV of Rhodobacter

capsulatus O2 sensitive without affecting its transductory activity. *FEBS Journal* **272**: 3899-3908.

Eguchi S, Yoon K-S, Ogo S. 2012. O 2-stable membrane-bound [NiFe] hydrogenase from a newly isolated Citrobacter sp. S-77. *Journal of bioscience and bioengineering* **114**: 479-484.

Ellis T, Wang X, Collins JJ. 2009. Diversity-based, model-guided construction of synthetic gene networks with predicted functions. *Nature biotechnology* **27**: 465-471.

Elowitz MB, Leibler S. 2000. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**: 335-338.

Evans RM, Brooke EJ, Wehlin SA, Nomerotskaia E, Sargent F, Carr SB, Phillips SE, Armstrong FA. 2016. Mechanism of hydrogen activation by [NiFe] hydrogenases. *Nature chemical biology* **12**: 46-50.

Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, Korlach J, Turner SW. 2010. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nature methods* **7**: 461-465.

Fontecilla-Camps JC, Volbeda A, Cavazza C, Nicolet Y. 2007. Structure/function relationships of [NiFe]-and [FeFe]-hydrogenases. *Chemical reviews* **107**: 4273-4303.

Forzi L, Sawers RG. 2007. Maturation of [NiFe]-hydrogenases in Escherichia coli. *Biometals* **20**: 565-578.

Friedland AE, Lu TK, Wang X, Shi D, Church G, Collins JJ. 2009. Synthetic gene networks that count. *Science* **324**: 1199-1202.

Friedrich B, Fritsch J, Lenz O. 2011. Oxygen-tolerant hydrogenases in hydrogen-based technologies. *Current opinion in biotechnology* **22**: 358-364.

Frielingsdorf S, Fritsch J, Schmidt A, Hammer M, Löwenstein J, Siebert E, Pelmenschikov V, Jaenicke T, Kalms J, Rippers Y. 2014. Reversible [4Fe-3S] cluster morphing in an O2-tolerant [NiFe] hydrogenase. *Nature chemical biology* **10**: 378-385.

Frielingsdorf S, Schubert T, Pohlmann A, Lenz O, Friedrich Br. 2011. A trimeric supercomplex of the oxygen-tolerant membrane-bound [NiFe]-hydrogenase from Ralstonia eutropha H16. *Biochemistry* **50**: 10836-10843.

Fritsch J, Lenz O, Friedrich B. 2013. Structure, function and biosynthesis of O2-tolerant hydrogenases. *Nature Reviews Microbiology* **11**: 106-114.

Fritsch J, Scheerer P, Frielingsdorf S, Kroschinsky S, Friedrich B, Lenz O, Spahn CM. 2011. The crystal structure of an oxygen-tolerant hydrogenase uncovers a novel iron-sulphur centre. *Nature* **479**: 249-252.

Gardner TS, Cantor CR, Collins JJ. 2000. Construction of a genetic toggle switch in Escherichia coli. *Nature* **403**: 339-342.

Garvie CW, Wolberger C. 2001. Recognition of specific DNA sequences. *Molecular cell* **8**: 937-946.

Gendrel A-V, Apedaile A, Coker H, Termanis A, Zvetkova I, Godwin J, Tang YA, Huntley D, Montana G, Taylor S. 2012. Smchd1-dependent and-independent pathways determine developmental dynamics of CpG island methylation on the inactive X chromosome. *Developmental cell* **23**: 265-279.

Gersbach CA, Perez-Pinera P, Black J. 2016. Mammalian Synthetic Biology: Opportunities and Prospects. *Annual review of biomedical engineering* **19**.

Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang R-Y, Algire MA, Benders GA, Montague MG, Ma L, Moodie MM. 2010a. Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329**: 52-56.

Gibson DG, Smith HO, Hutchison III CA, Venter JC, Merryman C. 2010b. Chemical synthesis of the mouse mitochondrial genome. *Nature methods* **7**: 901-903.

Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA, Smith HO. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature methods* **6**: 343-345.

Goedecke K, Pignot M, Goody RS, Scheidig AJ, Weinhold E. 2001. Structure of the N6-adenine DNA methyltransferase M• TaqI in complex with DNA and a cofactor analog. *Nature Structural & Molecular Biology* **8**: 121-125.

Gonzalez D, Collier J. 2013. DNA methylation by CcrM activates the transcription of two genes required for the division of Caulobacter crescentus. *Molecular microbiology* **88**: 203-218.

Gonzalez D, Kozdon JB, McAdams HH, Shapiro L, Collier J. 2014. The functions of DNA methylation by CcrM in Caulobacter crescentus: a global approach. *Nucleic acids research* **42**: 3720-3735.

Gorbatyuk B, Marczynski GT. 2005. Regulated degradation of chromosome replication proteins DnaA and CtrA in Caulobacter crescentus. *Molecular microbiology* **55**: 1233-1245.

Goris T, Wait AF, Saggu M, Fritsch J, Heidary N, Stein M, Zebger I, Lendzian F, Armstrong FA, Friedrich B. 2011. A unique iron-sulfur cluster is crucial for oxygen tolerance of a [NiFe]-hydrogenase. *Nature chemical biology* **7**: 310-318.

Green AA, Silver PA, Collins JJ, Yin P. 2014. Toehold switches: de-novo-designed regulators of gene expression. *Cell* **159**: 925-939.

Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT.

Hallenbeck PC, Abo-Hashesh M, Ghosh D. 2012. Strategies for improving biological hydrogen production. *Bioresource technology* **110**: 1-9.

Hallenbeck PC, Benemann JR. 2002. Biological hydrogen production; fundamentals and limiting processes. *international journal of hydrogen energy* **27**: 1185-1193.

Henikoff S, Greally JM. 2016. Epigenetics, cellular memory and gene regulation. *Current Biology* **26**: R644-R648.

Herrmann BG, Frischauf A-M. 1987. Isolation of genomic DNA. *Methods Enzymol;(United States)* **152**: 180-183.

Holtzendorff J, Hung D, Brende P, Reisenauer A, Viollier PH, McAdams HH, Shapiro L. 2004. Oscillating global regulators control the genetic circuit driving a bacterial cell cycle. *Science* **304**: 983-987.

Horsthemke B. 2014. In brief: genomic imprinting and imprinting diseases. *The Journal of pathology* **232**: 485-487.

Hu P, Ben-David Y, Milstein D. 2016. Rechargeable hydrogen storage system based on the dehydrogenative coupling of ethylenediamine with ethanol. *Angewandte Chemie International Edition* **55**: 1061-1064.

Hutchison CA, Chuang R-Y, Noskov VN, Assad-Garcia N, Deerinck TJ, Ellisman MH, Gill J, Kannan K, Karas BJ, Ma L. 2016. Design and synthesis of a minimal bacterial genome. *Science* **351**: aad6253.

Isaacs FJ, Dwyer DJ, Ding C, Pervouchine DD, Cantor CR, Collins JJ. 2004. Engineered riboregulators enable post-transcriptional control of gene expression. *Nature biotechnology* **22**: 841-847.

Jacob F, Monod J. 1961. On the regulation of gene activity. in *Cold Spring Harbor Symposia on Quantitative Biology*, pp. 193-211. Cold Spring Harbor Laboratory Press.

Jaenisch R, Bird A. 2003. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature genetics* **33**: 245-254.

Jeltsch A. 1999. Circular permutations in the molecular evolution of DNA methyltransferases. *Journal of molecular evolution* **49**: 161-164.

Jeltsch A. 2002. Beyond Watson and Crick: DNA methylation and molecular enzymology of DNA methyltransferases. *Chembiochem* **3**: 274-293.

Jeltsch A. 2006. On the enzymatic properties of Dnmt1: specificity, processivity, mechanism of linear diffusion and allosteric regulation of the enzyme. *Epigenetics* **1**: 63-66.

Jeltsch A, Friedrich T, Roth M. 1998. Kinetics of methylation and binding of DNA by the EcoRV adenine-N 6 methyltransferase. *Journal of molecular biology* **275**: 747-758.

Jeltsch A, Lanio T. 2002. Site-directed mutagenesis by polymerase chain reaction. *In Vitro Mutagenesis Protocols* **182**: 85-94.

Jonas K, Chen YE, Laub MT. 2011. Modularity of the bacterial cell cycle enables independent spatial and temporal control of DNA replication. *Current Biology* **21**: 1092-1101.

Jones PA. 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature Reviews Genetics* **13**: 484-492.

Jurkowski TP, Anspach N, Kulishova L, Nellen W, Jeltsch A. 2007. The M. EcoRV DNA-(adenine N6)-methyltransferase uses DNA bending for recognition of an expanded EcoDam recognition site. *Journal of Biological Chemistry* **282**: 36942-36952.

Jurkowski TP, Shanmugam R, Helm M, Jeltsch A. 2012. Mapping the tRNA binding site on the surface of human DNMT2 methyltransferase. *Biochemistry* **51**: 4438-4444.

Kahng LS, Shapiro L. 2001. The CcrM DNA methyltransferase of Agrobacterium tumefaciens is essential, and its activity is cell cycle regulated. *Journal of bacteriology* **183**: 3065-3075.

Kapdan IK, Kargi F. 2006. Bio-hydrogen production from waste materials. *Enzyme and microbial technology* **38**: 569-582.

Keasling JD. 2010. Manufacturing molecules through metabolic engineering. *Science* **330**: 1355-1358.

Khalil AS, Collins JJ. 2010. Synthetic biology: applications come of age. *Nature Reviews Genetics* **11**: 367-379.

Klemm P. 1986. Two regulatory fim genes, fimB and fimE, control the phase variation of type 1 fimbriae in Escherichia coli. *The EMBO journal* **5**: 1389.

Kobayashi H, Kærn M, Araki M, Chung K, Gardner TS, Cantor CR, Collins JJ. 2004. Programmable cells: interfacing natural and engineered gene networks. *Proceedings of the National Academy of Sciences of the United States of America* **101**: 8414-8419.

Konermann S, Brigham MD, Trevino AE, Joung J, Abudayyeh OO, Barcena C, Hsu PD, Habib N, Gootenberg JS, Nishimasu H. 2014. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**: 583-588.

Korpela MT, Kurittu JS, Karvinen JT, Karp MT. 1998. A recombinant Escherichia coli sensor strain for the detection of tetracyclines. *Analytical chemistry* **70**: 4457-4462.

Kothandaraman J, Kar S, Sen R, Goeppert A, Olah GA, Prakash GS. 2017. An Efficient Reversible Hydrogen Carrier System Based on Amine Reforming of Methanol. *Journal of the American Chemical Society* **139**: 2549-2552.

Kothari R, Singh D, Tyagi V, Tyagi S. 2012. Fermentative hydrogen production–An alternative clean energy source. *Renewable and Sustainable Energy Reviews* **16**: 2337-2346.

Kozdon JB, Melfi MD, Luong K, Clark TA, Boitano M, Wang S, Zhou B, Gonzalez D, Collier J, Turner SW. 2013. Global methylation state at base-pair resolution of the Caulobacter genome throughout the cell cycle. *Proceedings of the National Academy of Sciences* **110**: E4658-E4667.

Kröger A, Geisler V, Lemma E, Theis F, Lenger R. 1992. Bacterial fumarate respiration. *Archives of microbiology* **158**: 311-314.

Labahn J, Granzin J, Schluckebier G, Robinson D, Jack W, Schildkraut I, Saenger W. 1994. Three-dimensional structure of the adenine-specific DNA methyltransferase M. Taq I in complex with the cofactor S-adenosylmethionine. *Proceedings of the National Academy of Sciences* **91**: 10957-10961.

Laub MT, Chen SL, Shapiro L, McAdams HH. 2002. Genes directly controlled by CtrA, a master regulator of the Caulobacter cell cycle. *Proceedings of the National Academy of Sciences* **99**: 4632-4637.

Lee JW, Na D, Park JM, Lee J, Choi S, Lee SY. 2012. Systems metabolic engineering of microorganisms for natural and non-natural chemicals. *Nature chemical biology* **8**: 536-546.

Lee S, Doddapaneni K, Hogue A, McGhee L, Meyers S, Wu Z. 2010. Solution structure of Gfi-1 zinc domain bound to consensus DNA. *Journal of molecular biology* **397**: 1055-1066.

Lenz O, Ludwig M, Schubert T, Bürstel I, Ganskow S, Goris T, Schwarze A, Friedrich B. 2010. H2 Conversion in the Presence of O2 as Performed by the Membrane-Bound [NiFe]-Hydrogenase of Ralstonia eutropha. *ChemPhysChem* **11**: 1107-1119.

Leonhartsberger S, Korsa I, Bock A. 2002. The molecular biology of formate metabolism in enterobacteria. *Journal of molecular microbiology and biotechnology* **4**: 269-276.

Li G-W, Burkhardt D, Gross C, Weissman JS. 2014. Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell* **157**: 624-635.

Lienert F, Lohmueller JJ, Garg A, Silver PA. 2014. Synthetic biology in mammalian cells: next generation research tools and therapeutics. *Nature reviews Molecular cell biology* **15**: 95-107.

Liu Y, Olanrewaju YO, Zheng Y, Hashimoto H, Blumenthal RM, Zhang X, Cheng X. 2014. Structural basis for Klf4 recognition of methylated DNA. *Nucleic acids research* **42**: 4859-4867.

Liu Y, Toh H, Sasaki H, Zhang X, Cheng X. 2012. An atomic model of Zfp57 recognition of CpG methylation within a specific DNA sequence. *Genes & development* **26**: 2374-2379.

Løbner-Olesen A, Skovgaard O, Marinus MG. 2005. Dam methylation: coordinating cellular processes. *Current opinion in microbiology* **8**: 154-160.

Low DA, Weyand NJ, Mahan MJ. 2001. Roles of DNA adenine methylation in regulating bacterial gene expression and virulence. *Infection and Immunity* **69**: 7197-7204.

Lubitz W, Ogata H, Rüdiger O, Reijerse E. 2014. Hydrogenases. *Chemical reviews* **114**: 4081-4148.

Lukey MJ, Parkin A, Roessler MM, Murphy BJ, Harmer J, Palmer T, Sargent F, Armstrong FA. 2010. How Escherichia coli is equipped to oxidize hydrogen under different redox conditions. *Journal of Biological Chemistry* **285**: 3928-3938.

Lukey MJ, Roessler MM, Parkin A, Evans RM, Davies RA, Lenz O, Friedrich B, Sargent F, Armstrong FA. 2011. Oxygen-tolerant [NiFe]-hydrogenases: the individual and collective importance of supernumerary cysteines at the proximal Fe-S cluster. *Journal of the American Chemical Society* **133**: 16881-16892.

Luo G-Z, Blanco MA, Greer EL, He C, Shi Y. 2015. DNA N6-methyladenine: a new epigenetic mark in eukaryotes? *Nature reviews Molecular cell biology* **16**: 705-710.

Luscombe NM, Laskowski RA, Thornton JM. 2001. Amino acid–base interactions: a three-dimensional analysis of protein–DNA interactions at an atomic level. *Nucleic acids research* **29**: 2860-2874.

Lutz R, Bujard H. 1997. Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic acids research* **25**: 1203-1210.

Maier JA, Albu RF, Jurkowski TP, Jeltsch A. 2015a. Investigation of the C-terminal domain of the bacterial DNA-(adenine N6)-methyltransferase CcrM. *Biochimie* **119**: 60-67.

Maier JA, Ragozin S, Jeltsch A. 2015b. Identification, cloning and heterologous expression of active [NiFe]-hydrogenase 2 from Citrobacter sp. SG in Escherichia coli. *Journal of biotechnology* **199**: 1-8.

Maier JAH, Möhrle R, Jeltsch A. 2017. Design of synthetic epigenetic circuits featuring memory effects and reversible switching based on DNA methylation. *Nature Communications* **8**: 15336.

Malone T, Blumenthal RM, Cheng X. 1995. Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyl-transferases, and suggests a catalytic mechanism for these enzymes. *Journal of molecular biology* **253**: 618-632.

Mandell DJ, Lajoie MJ, Mee MT, Takeuchi R, Kuznetsov G, Norville JE, Gregg CJ, Stoddard BL, Church GM. 2015. Biocontainment of genetically modified organisms by synthetic protein design. *Nature* **518**: 55-60.

Marczynski GT. 1999. Chromosome Methylation and Measurement of Faithful, Once and Only Once per Cell Cycle Chromosome Replication inCaulobacter crescentus. *Journal of bacteriology* **181**: 1984-1993.

Marczynski GT, Shapiro L. 2002. Control of chromosome replication in Caulobacter crescentus. *Annual Reviews in Microbiology* **56**: 625-656.

Marinus MG, Casadesus J. 2009. Roles of DNA adenine methylation in host–pathogen interactions: mismatch repair, transcriptional regulation, and more. *FEMS microbiology reviews* **33**: 488-503.

Marinus MG, Morris NR. 1973. Isolation of deoxyribonucleic acid methylase mutants of Escherichia coli K-12. *Journal of bacteriology* **114**: 1143-1150.

Matsumoto T, Eguchi S, Nakai H, Hibino T, Yoon KS, Ogo S. 2014. [NiFe] Hydrogenase from Citrobacter sp. S-77 Surpasses Platinum as an Electrode for H2 Oxidation Reaction. *Angewandte Chemie International Edition* **53**: 8895-8898.

Meinhardt S, Manley MW, Becker NA, Hessman JA, Maher LJ, Swint-Kruse L. 2012. Novel insights from hybrid LacI/GalR proteins: family-wide functional attributes and biologically significant variation in transcription repression. *Nucleic acids research* **40**: gks806.

Meinhardt S, Swint-Kruse L. 2008. Experimental identification of specificity determinants in the domain linker of a LacI/GalR protein: Bioinformatics-based

predictions generate true positives and false negatives. *Proteins: Structure, Function, and Bioinformatics* **73**: 941-957.

Meyer J, Kelley BC, Vignais PM. 1978. Aerobic nitrogen fixation by Rhodopseudomonas capsulata. *FEBS letters* **85**: 224-228.

Mohapatra SS, Fioravanti A, Biondi EG. 2014. DNA methylation in Caulobacter and other Alphaproteobacteria during cell cycle progression. *Trends in microbiology* **22**: 528-535.

Monod J, Jacob F. 1961. General conclusions: teleonomic mechanisms in cellular metabolism, growth, and differentiation. in *Cold Spring Harbor symposia on quantitative biology*, pp. 389-401. Cold Spring Harbor Laboratory Press.

Muhd Noor N, Nishikawa K, Nishihara H, Yoon K-S, Ogo S, Higuchi Y. 2016. Improved purification, crystallization and crystallographic study of Hyd-2-type [NiFe]-hydrogenase from Citrobacter sp. S-77. *Acta Crystallographica Section F: Structural Biology Communications* **72**: 53-58.

Nemhauser JL, Torii KU. 2016. Plant synthetic biology for molecular engineering of signalling and development. *Nature plants* **2**: 16010.

Nephew KP, Balch C, Skalnik DG. 2009. Methyl group acceptance assay for the determination of global DNA methylation levels. *DNA Methylation: Methods and Protocols* **507**: 35-41.

Nield J, Kruse O, Ruprecht J, da Fonseca P, Büchel C, Barber J. 2000. Three-dimensional structure of Chlamydomonas reinhardtii and Synechococcus elongatus photosystem II complexes allows for comparison of their oxygen-evolving complex organization. *Journal of Biological Chemistry* **275**: 27940-27946.

Nielsen AA, Der BS, Shin J, Vaidyanathan P, Paralanov V, Strychalski EA, Ross D, Densmore D, Voigt CA. 2016. Genetic circuit design automation. *Science* **352**: aac7341.

Nishimasu H, Ran FA, Hsu PD, Konermann S, Shehata SI, Dohmae N, Ishitani R, Zhang F, Nureki O. 2014. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**: 935-949.

Okano M, Bell DW, Haber DA, Li E. 1999. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**: 247-257.

Olson EJ, Tabor JJ. 2012. Post-translational tools expand the scope of synthetic biology. *Current opinion in chemical biology* **16**: 300-306.

Oppenheim AB, Kobiler O, Stavans J, Court DL, Adhya S. 2005. Switches in bacteriophage lambda development. *Annu Rev Genet* **39**: 409-429.

Otwinowski Z, Schevitz R, Zhang R, Lawson C, Joachimiak A, Marmorstein R, Luisi B, Sigler P. 1988. Crystal structure of trp represser/operator complex at atomic resolution. *Nature* **335**: 321-329.

Panis G, Murray SR, Viollier PH. 2014. Versatility of global transcriptional regulators in alpha-Proteobacteria: from essential cell cycle control to ancillary functions. *FEMS microbiology reviews* **39**: fuu002.

Pingoud A, Jeltsch A. 2001. Structure and function of type II restriction endonucleases. *Nucleic acids research* **29**: 3705-3727.

Pinske C, Jaroschinsky M, Linek S, Kelly CL, Sargent F, Sawers RG. 2015. Physiology and bioenergetics of [NiFe]-hydrogenase 2-catalyzed H2-consuming and H2-producing reactions in Escherichia coli. *Journal of bacteriology* **197**: 296-306.

Ptashne M. 2004. *A genetic switch: phage lambda revisited*. Cold Spring Harbor Laboratory Press Cold Spring Harbor, NY:.

Qi LS, Arkin AP. 2014. A versatile framework for microbial engineering using synthetic non-coding RNAs. *Nature Reviews Microbiology* **12**: 341-354.

Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA. 2013. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**: 1173-1183.

Rand AC, Jain M, Eizenga JM, Musselman-Brown A, Olsen HE, Akeson M, Paten B. 2017. Mapping DNA methylation with high-throughput nanopore sequencing. *Nature methods*.

Reich NO, Mashhoon N. 1991. Kinetic mechanism of the EcoRI DNA methyltransferase. *Biochemistry* **30**: 2933-2939.

Reisenauer A, Kahng LS, McCollum S, Shapiro L. 1999a. Bacterial DNA methylation: a cell cycle regulator? *Journal of bacteriology* **181**: 5135-5139.

Reisenauer A, Quon K, Shapiro L. 1999b. The CtrA response regulator mediates temporal control of gene expression during the Caulobacter cell cycle. *Journal of bacteriology* **181**: 2430-2439.

Reisenauer A, Shapiro L. 2002. DNA methylation affects the cell cycle transcription of the CtrA global regulator in Caulobacter. *The EMBO journal* **21**: 4969-4977.

Richard DJ, Sawers G, Sargent F, McWalter L, Boxer DH. 1999. Transcriptional regulation in response to oxygen and nitrate of the operons encoding the [NiFe] hydrogenases 1 and 2 of Escherichia coli. *Microbiology* **145**: 2903-2912.

Ro D-K, Paradise EM, Ouellet M, Fisher KJ, Newman KL, Ndungu JM, Ho KA, Eachus RA, Ham TS, Kirby J. 2006. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**: 940-943.

Roberts RC, Shapiro L. 1997. Transcription of genes encoding DNA replication proteins is coincident with cell cycle control of DNA replication in Caulobacter crescentus. *Journal of bacteriology* **179**: 2319-2330.

Robertson GT, Reisenauer A, Wright R, Jensen RB, Jensen A, Shapiro L, Roop RM. 2000. The Brucella abortus CcrM DNA methyltransferase is essential for viability, and its overexpression attenuates intracellular replication in murine macrophages. *Journal of bacteriology* **182**: 3482-3489.

Rohs R, Jin X, West SM, Joshi R, Honig B, Mann RS. 2010. Origins of specificity in protein-DNA recognition. *Annual review of biochemistry* **79**: 233.

Roquet N, Soleimany AP, Ferris AC, Aaronson S, Lu TK. 2016. Synthetic recombinase-based state machines in living cells. *Science* **353**: aad8559.

Roth M, Jeltsch A. 2000. Biotin-avidin microplate assay for the quantitative analysis of enzymatic methylation of DNA by DNA methyltransferases. *Biological chemistry* **381**: 269-272.

Rovner AJ, Haimovich AD, Katz SR, Li Z, Grome MW, Gassaway BM, Amiram M, Patel JR, Gallagher RR, Rinehart J. 2015. Recoded organisms engineered to depend on synthetic amino acids. *Nature* **518**: 89-93.

Sasai N, Nakao M, Defossez P-A. 2010. Sequence-specific recognition of methylated DNA by human zinc-finger proteins. *Nucleic acids research* **38**: 5015-5022.

Sawers G. 1994. The hydrogenases and formate dehydrogenases ofEscherichia coli. *Antonie Van Leeuwenhoek* **66**: 57-88.

Schlapbach L, Züttel A. 2001. Hydrogen-storage materials for mobile applications. *Nature* **414**: 353-358.

Schluckebier G, Kozak M, Bleimling N, Weinhold E, Saenger W. 1997. Differential binding of S-adenosylmethionine S-adenosylhomocysteine and Sinefungin to the adenine-specific DNA methyltransferase M. TaqI. *Journal of molecular biology* **265**: 56-67.

Schluckebier G, O'Gara M, Saenger W, Cheng X. 1995. Universal catalytic domain structure of AdoMet-dependent methyltransferases. *Journal of molecular biology* **247**: 16-20.

Seeman NC, Rosenberg JM, Rich A. 1976. Sequence-specific recognition of double helical nucleic acids by proteins. *Proceedings of the National Academy of Sciences* **73**: 804-808.

Sera T, Uranga C. 2002. Rational design of artificial zinc-finger proteins using a nondegenerate recognition code table. *Biochemistry* **41**: 7074-7081.

Serban M, Lewis MA, Marshall CL, Doctor RD. 2003. Hydrogen production by direct contact pyrolysis of natural gas. *Energy & fuels* **17**: 705-713.

Serganov A, Nudler E. 2013. A decade of riboswitches. *Cell* **152**: 17-24.

Sethia G, Sayari A. 2016. Activated carbon with optimum pore size distribution for hydrogen storage. *Carbon* **99**: 289-294.

Shier VK, Hancey CJ, Benkovic SJ. 2001. Identification of the active oligomeric state of an essential adenine DNA methyltransferase from Caulobacter crescentus. *Journal of Biological Chemistry* **276**: 14744-14751.

Shomura Y, Yoon K-S, Nishihara H, Higuchi Y. 2011. Structural basis for a [lsqb] 4Fe-3S [rsqb] cluster in the oxygen-tolerant membrane-bound [lsqb] NiFe [rsqb]-hydrogenase. *Nature* **479**: 253-256.

Simon E, Kornitzer D. 2014. Pulse-chase analysis to measure protein degradation. *Methods Enzymol* **536**: 65-75.

Simpson JT, Workman RE, Zuzarte P, David M, Dursi L, Timp W. 2017. Detecting DNA cytosine methylation using nanopore sequencing. *Nature methods*.

Siuti P, Yazbek J, Lu TK. 2013. Synthetic circuits integrating logic and memory in living cells. *Nature biotechnology* **31**: 448-452.

Skarstad K, Boye E. 1994. The initiator protein DnaA: evolution, properties and function. *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression* **1217**: 111-130.

Skerker JM, Perchuk BS, Siryaporn A, Lubin EA, Ashenberg O, Goulian M, Laub MT. 2008. Rewiring the specificity of two-component signal transduction systems. *Cell* **133**: 1043-1054.

Smanski MJ, Zhou H, Claesen J, Shen B, Fischbach MA, Voigt CA. 2016. Synthetic biology to access and expand nature's chemical diversity. *Nature Reviews Microbiology* **14**: 135-149.

Smith ZD, Meissner A. 2013. DNA methylation: roles in mammalian development. *Nature Reviews Genetics* **14**: 204-220.

Stanton BC, Nielsen AA, Tamsir A, Clancy K, Peterson T, Voigt CA. 2014. Genomic mining of prokaryotic repressors for orthogonal logic gates. *Nature chemical biology* **10**: 99-105.

Stripp ST, Goldet G, Brandmayr C, Sanganas O, Vincent KA, Haumann M, Armstrong FA, Happe T. 2009. How oxygen attacks [FeFe] hydrogenases from photosynthetic organisms. *Proceedings of the National Academy of Sciences* **106**: 17331-17336.

Sugimoto K, Takeda S, Hirochika H. 2003. Transcriptional activation mediated by binding of a plant GATA-type zinc finger protein AGP1 to the AG-motif (AGATCCAA) of the wound-inducible Myb gene NtMyb2. *The Plant Journal* **36**: 550-564.

Suzuki MM, Bird A. 2008. DNA methylation landscapes: provocative insights from epigenomics. *Nature Reviews Genetics* **9**: 465-476.

Tamsir A, Tabor JJ, Voigt CA. 2011. Robust multicellular computing using genetically encoded NOR gates and chemical/wires/'. *Nature* **469**: 212-215.

Taylor JD, Badcoe IG, Clarke AR, Halford SE. 1991. EcoRV restriction endonuclease binds all DNA sequences with equal affinity. *Biochemistry* **30**: 8743-8753.

Thauer RK, Kaster A-K, Goenrich M, Schick M, Hiromoto T, Shima S. 2010. Hydrogenases from methanogenic archaea, nickel, a novel cofactor, and H2 storage. *Annual review of biochemistry* **79**: 507-536.

Thompson JD, Gibson T, Higgins DG. 2002. Multiple sequence alignment using ClustalW and ClustalX. *Current protocols in bioinformatics*: 2.3. 1-2.3. 22.

Travers AA. 1989. DNA conformation and protein binding. *Annual review of biochemistry* **58**: 427-452.

van der Woude M, Braaten B, Low D. 1996. Epigenetic phase variation of the pap operon in Escherichia coli. *Trends in microbiology* **4**: 5-9.

Vignais PM, Billoud B. 2007. Occurrence, classification, and biological function of hydrogenases: an overview. *Chemical reviews* **107**: 4206-4272.

Vincent KA, Parkin A, Armstrong FA. 2007. Investigating and exploiting the electrocatalytic properties of hydrogenases. *Chemical reviews* **107**: 4366-4413.

Vincent KA, Parkin A, Lenz O, Albracht SP, Fontecilla-Camps JC, Cammack R, Friedrich B, Armstrong FA. 2005. Electrochemical definitions of O2 sensitivity and oxidative inactivation in hydrogenases. *Journal of the American Chemical Society* **127**: 18179-18189.

Volbeda A, Amara P, Darnault C, Mouesca J-M, Parkin A, Roessler MM, Armstrong FA, Fontecilla-Camps JC. 2012. X-ray crystallographic and computational studies of the O2-tolerant [NiFe]-hydrogenase 1 from Escherichia coli. *Proceedings of the National Academy of Sciences* **109**: 5305-5310.

Wang B, Kitney RI, Joly N, Buck M. 2011. Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology. *Nature communications* **2**: 508.

Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G, Forest CR, Church GM. 2009a. Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**: 894-898.

Wang M, Wang Z, Gong X, Guo Z. 2014. The intensification technologies to water electrolysis for hydrogen production–A review. *Renewable and Sustainable Energy Reviews* **29**: 573-588.

Wang X, Maeda K, Thomas A, Takanabe K, Xin G, Carlsson JM, Domen K, Antonietti M. 2009b. A metal-free polymeric photocatalyst for hydrogen production from water under visible light. *Nature materials* **8**: 76-80.

Watanabe S, Sasaki D, Tominaga T, Miki K. 2012. Structural basis of [NiFe] hydrogenase maturation by Hyp proteins. *Biol Chem* **393**: 1089-1100.

Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. 2009. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**: 1189-1191.

Waters LS, Storz G. 2009. Regulatory RNAs in bacteria. *Cell* **136**: 615-628.

Weber W, Fussenegger M. 2012. Emerging biomedical applications of synthetic biology. *Nature Reviews Genetics* **13**: 21-35.

Wion D, Casadesús J. 2006. N6-methyl-adenine: an epigenetic signal for DNA–protein interactions. *Nature Reviews Microbiology* **4**: 183-192.

Wolfe SA, Nekludova L, Pabo CO. 2000. DNA recognition by Cys2His2 zinc finger proteins. *Annual review of biophysics and biomolecular structure* **29**: 183-212.

Wright DA, Thibodeau-Beganny S, Sander JD, Winfrey RJ, Hirsh AS, Eichtinger M, Fu F, Porteus MH, Dobbs D, Voytas DF. 2006. Standardized reagents and protocols for engineering zinc finger nucleases by modular assembly. *Nature protocols* **1**: 1637-1652.

Wright R, Stephens C, Zweiger G, Shapiro L, Alley M. 1996. Caulobacter Lon protease has a critical role in cell-cycle control of DNA methylation. *Genes & development* **10**: 1532-1542.

Wu H, Song Z, Hentzer M, Andersen JB, Heydorn A, Mathee K, Moser C, Eberl L, Molin S, Høiby N. 2000. Detection of N-acylhomoserine lactones in lung tissues of mice infected with Pseudomonas aeruginosa. *Microbiology* **146**: 2481-2493.

Wu TP, Wang T, Seetin MG, Lai Y, Zhu S, Lin K, Liu Y, Byrum SD, Mackintosh SG, Zhong M. 2016. DNA methylation on N6-adenine in mammalian embryonic stem cells. *Nature*.

Wulff P, Thomas C, Sargent F, Armstrong FA. 2016. How the oxygen tolerance of a [NiFe]-hydrogenase depends on quaternary structure. *JBIC Journal of Biological Inorganic Chemistry* **21**: 121-134.

Yang L, Nielsen AA, Fernandez-Rodriguez J, McClune CJ, Laub MT, Lu TK, Voigt CA. 2014. Permanent genetic memory with> 1-byte capacity. *Nature methods* **11**: 1261-1266.

Zhan J, Ding B, Ma R, Ma X, Su X, Zhao Y, Liu Z, Wu J, Liu H. 2010. Develop reusable and combinable designs for transcriptional logic gates. *Molecular systems biology* **6**: 388.

Zhang Y, Xi Z, Hegde RS, Shakked Z, Crothers DM. 2004. Predicting indirect readout effects in protein–DNA interactions. *Proceedings of the National Academy of Sciences of the United States of America* **101**: 8337-8341.

Zweiger G, Marczynski G, Shapiro L. 1994. A Caulobacter DNA methyltransferase that functions only in the predivisional cell. *Journal of molecular biology* **235**: 472-485.

Author’s contribution

Maier, J.A.H., Möhrle, R., Jeltsch, A. **Design of synthetic epigenetic circuits featuring memory effects and reversible switching based on DNA methylation**. Nature Communications 8: 15336. (2017)

JM and AJ designed the study. JM carried out all experiments. RM contributed to cloning, expression and purification of zinc finger protein, EMSA experiments and memory system IV (reset switch). JM and AJ analyzed and interpreted the data. JM and AJ wrote the manuscript draft.

Maier, J.A.H., Albu, R.F., Jurkowski, T.P. & Jeltsch, A. **Investigation of the C-terminal domain of the bacterial DNA-(adenine N6)-methyltransferase CcrM.** Biochimie 119, 60-67 (2015).

All authors contributed to the design of the study. RA conducted the experiments shown in Fig. 4A and B and Supplementary Fig. 2, JM conducted all other experiments. All authors were involved in data analysis and interpretation. JM and AJ wrote the manuscript draft.

Maier, J.A.H., Ragozin, S. & Jeltsch, A. **Identification, cloning and heterologous expression of active [NiFe]-hydrogenase 2 from Citrobacter sp. SG in Escherichia coli.** Journal of biotechnology 199, 1-8 (2015).

JM participated in the design of the study and carried out all experiments. JM and AJ wrote the manuscript. Characterization and identification of the hydrogen producing bacterium *Citrobacter* sp. was conducted in scope of my diploma thesis.

Appendix

Appendix 1 (not included in the published thesis)

Maier, J.A.H., Albu, R.F., Jurkowski, T.P. & Jeltsch, A. Investigation of the C-terminal domain of the bacterial DNA-(adenine N6)-methyltransferase CcrM. Biochimie 119, 60-67 (2015).

Appendix 2 (not included in the published thesis)

Maier J.A.H., Möhrle R, Jeltsch A. Design of synthetic epigenetic circuits featuring memory effects and reversible switching based on DNA methylation. Nature Communications 8: 15336. (2017)

Appendix 3 (not included in the published thesis)

Maier, J.A.H., Ragozin, S. & Jeltsch, A. Identification, cloning and heterologous expression of active [NiFe]-hydrogenase 2 from Citrobacter sp. SG in Escherichia coli. Journal of biotechnology 199, 1-8 (2015).

Appendix 4

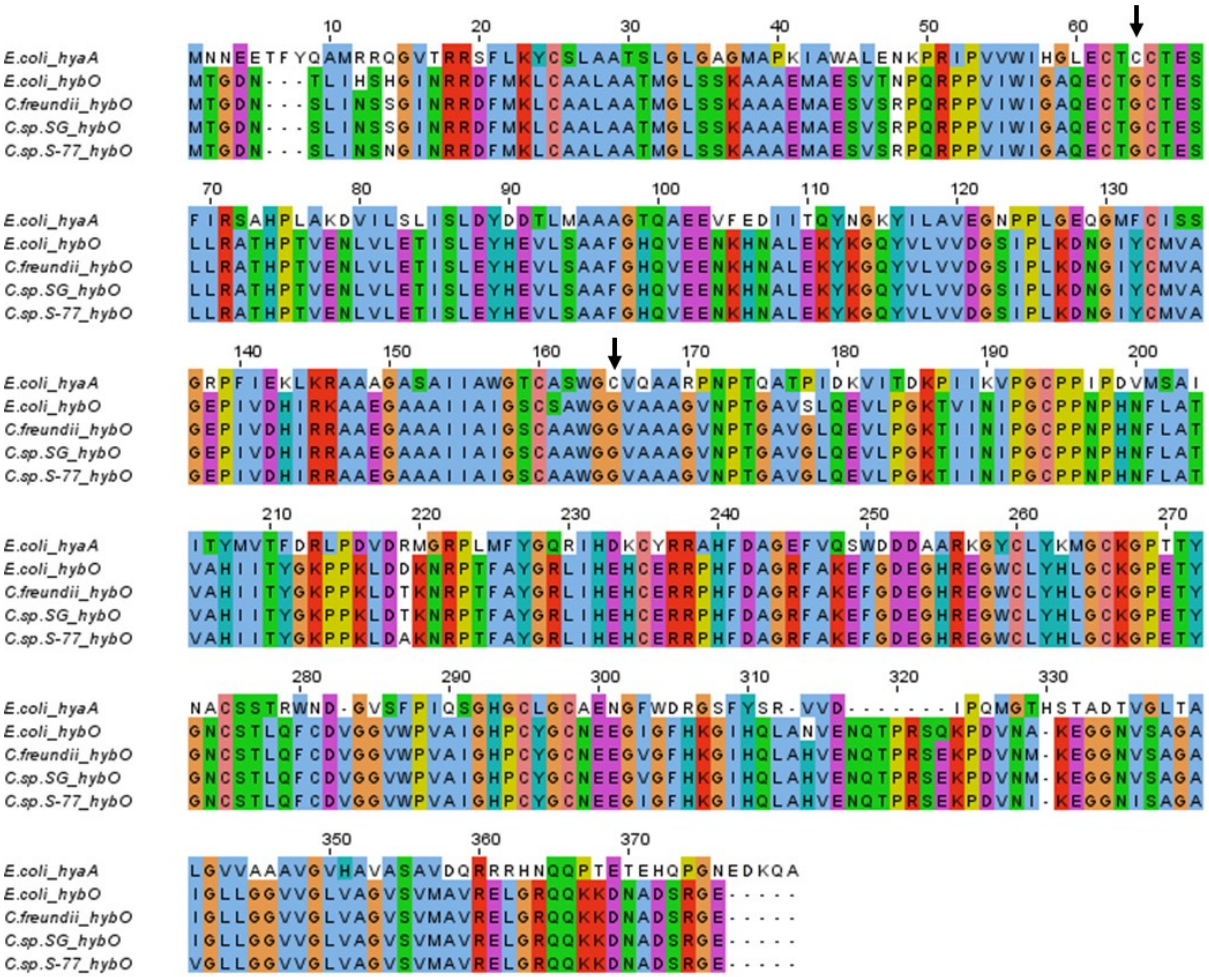
Sequence alignment hydrogenases small subunits

Appendix 5

Sequence alignment hydrogenases large subunits

Appendix 4

Multiple sequence alignment of small subunits of hydrogenase 1 and 2 from *Escherichia coli* and small subunits of hydrogenases 2 from *Citrobacter freundii*, *Citrobacter* sp. SG, *Citrobacter* sp S-77. Indicated by the black arrows are cysteine residues responsible for coordination of [4Fe3S] cluster in oxygen tolerant hydrogenase 1 and not present in type 2 variants.



Appendix 5

Multiple sequence alignment of large subunits of hydrogenase 1 and 2 from *Escherichia coli* and large subunits of hydrogenases 2 from *Citrobacter freundii*, *Citrobacter* sp. SG, *Citrobacter* sp S-77. Indicated by the black arrows are positions of cysteine residues present only in hydrogenase 2 large subunits of Citrobacter species.