Uncovering the transcriptional control of *Bartonella henselae* host adaptation factors

Inaugural dissertation

zur
Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel

von

Maxime Québatte
aus Saignelégier, Schweiz

Basel, 2014

Original document stored on the publication server of the University of Basel
edoc.unibas.ch

This work is licenced under the agreement „Attribution Non-Commercial No Derivatives – 3.0 Switzerland“ (CC BY-NC-ND 3.0 CH).
The complete text may be reviewed here:
creativecommons.org/licenses/by-nc-nd/3.0/ch/deed.en
Attribution-NonCommercial-NoDerivatives 3.0 Switzerland
(CC BY-NC-ND 3.0 CH)

You are free: to Share — to copy, distribute and transmit the work

Under the following conditions:

Attribution — You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).

Noncommercial — You may not use this work for commercial purposes.

No Derivative Works — You may not alter, transform, or build upon this work.

With the understanding that:

• Waiver — Any of the above conditions can be waived if you get permission from the copyright holder.

• Public Domain — Where the work or any of its elements is in the public domain under applicable law, that status is in no way affected by the license.

• Other Rights — In no way are any of the following rights affected by the license:
  o Your fair dealing or fair use rights, or other applicable copyright exceptions and limitations;
  o The author's moral rights;
  o Rights other persons may have either in the work itself or in how the work is used, such as publicity or privacy rights.

• Notice — For any reuse or distribution, you must make clear to others the license terms of this work. The best way to do this is with a link to this web page.

Quelle:creativecommons.org/licenses/by-nc-nd/3.0/ch/deed.en                  Datum: 12.11.2013
“Not only is the Universe stranger than we think, it is stranger than we can think.”

Werner Heisenberg, *Across the Frontiers*
STATEMENT OF MY THESIS

This work was performed in the group of Prof. Christoph Dehio in the Focal Area Infection Biology at the Biozentrum of the University of Basel. My PhD advisory Committee consisted of:

Prof. Christoph Dehio
Prof. Urs Jenal
Prof. Dirk Bumann

My thesis is written in a cumulative format. It consists of an abstract, a synopsis covering several aspects related to my work, a result section composed of four scientific publications and an unpublished results section followed by some concluding remarks.
Abstract

A recurrent theme in bacterial pathogenicity is the understanding of the regulatory events necessary for a given pathogen to progress through its infection cycle while resisting the host defense mechanisms. This progression typically requires the coordinated expression of defined sub-portions of the virulence repertoire at the same time as others need to be tightly repressed or degraded. This so-called adaptive response is ultimately linked to the ability of the pathogen to sense its direct environment and to transduce this information into the appropriate cellular response. Bacteria have evolved numerous dedicated mechanisms for perception and signal transduction that are characterized by a wide range of signal specificity. Not surprisingly, most of these systems have been adopted by pathogenic bacteria to modulate the expression of their virulence factors. In this work, we present the results of our investigations on the mounting and the regulation of the adaptive response of the zoonotic bacterial pathogen *Bartonella henselae* to its eukaryotic host. The VirB/D4 type IV secretion system (T4SS) is an essential machinery for the host adaptation of this stealthy pathogen. Using the regulation of this pathogenicity factor as a red thread we uncovered two critical signal transductions pathways that enable *B. henselae* to coordinate the expression of its virulence factors through its infection cycle.

In the *research article I*, we describe the adaptive response of *B. henselae* during host cell infection and reveal the central role of the BatR/BatS two component system (TCS) for the coordination of this response. We demonstrate that this TCS is activated at the physiological pH of blood (pH7.4) and is required for the up-regulation of a critical cluster of genes that includes the genes encoding the VirB/D4 T4SS and its cognate secreted effectors (Beps). In the *research article II*, we present the near complete expressed proteome of *B. henselae* under conditions that mimic host-interaction, using a combination of saturated transcriptome profiling by RNA-seq and directed shotgun proteomics. Of particular interest, the complete membrane proteome coverage achieved reveals the dramatic re-organization taking place in this compartment during the infection process, with the differential regulation of a large panel of autotransporters, adhesins and hemin binding proteins as well as all components of the VirB/D4 T4SS. In the *research article III*, we describe how a dual regulatory input controls the expression of the VirB/D4 T4SS and its secreted effector proteins. We demonstrate that additionally to the
BatR/BatS TCS, the expression of this host adaptation factors requires the alternative sigma factor RpoH1, which is itself controlled by the stringent response (SR) components SpoT and DksA. In contrast to the VirB/D4 T4SS, which is needed at the early stage of mammalian host colonization and require the SR components for its full expression, we show that SpoT and DksA negatively regulate the Trw T4SS, which is required for erythrocyte invasion at a later stage of the host infection. In the research article IV, we demonstrate the possible use of B. henselae to deliver DNA into human cells through its VirB/D4 T4SS and to generate stable transgenic cell lines. We propose that due to its ancestral abilities as conjugation system, this specialized transkingdom secretion system has potential for the development of new in vivo gene therapy approaches in humans.

Together, these results constitute the first comprehensive analysis of B. henselae pathogenicity factors during host cell infection. Besides the elucidation of very specific regulatory aspects for the expression of the VirB/D4 T4SS and its secreted substrates, this work allows us to propose a general model for B. henselae host adaptation strategy throughout its infection cycle. In our model, the BatR/BatS pH-dependent signaling is used to distinguish between the arthropod and the mammalian host environment whereas the SR signaling allows the modulation of the bacterial response between the early and the late colonization stages of the mammalian host.
TABLE OF CONTENT

1. INTRODUCTION .............................................................. 1

1.1. BACKGROUND .......................................................... 3

1.2. SIGNAL TRANSDUCTION MODULES ......................... 3

1.3. THE PARADIGM OF BACTERIAL TCSs ....................... 4

1.3.1. Domain organization .............................................. 5

1.3.2. General signal transduction mechanism .................... 7

1.3.3. Activation of HKs upon signal recognition ................ 7

1.3.4. Activation of RRs by HKs ....................................... 9

1.3.5. Activation of the OmpR/PhoB subfamily .................... 9

1.3.6. DNA binding by the OmpR/PhoB family of RRs .......... 10

1.3.7. Transcription activation by RRs ............................... 12

1.4. SPECIFICITY IN TCS SIGNALING .............................. 13

1.5. MODULATION OF TCSs BY ACCESSORY PROTEINS .... 14

1.6. GENOME-WIDE APPROACHES TO STUDY TCSs .......... 16

1.6.1. Genome wide transcription analysis applied to TCS signaling .......... 16

1.6.2. Genome wide mapping of RRs DNA binding sites .......... 18

1.7. TCSs AND TRANSCRIPTION NETWORKS ..................... 20

1.7.1. Positive feedback regulation of TCSs ........................ 22

1.7.2. Bistability in the context of TCS regulation ............... 24

1.7.3. Global transcription networks and TCSs ..................... 25

1.8. REFERENCES ........................................................... 27

2. AIM OF THE THESIS .................................................. 35

3. RESULTS ................................................................. 39

3.1. RESEARCH ARTICLE I .............................................. 41

3.1.1. Statement of the own participation ........................... 41


3.2. RESEARCH ARTICLE II ............................................ 59

3.2.1. Statement of the own participation ........................... 59

3.2.2. Manuscript: “Directed shotgun proteomics guided by saturated RNA-seq identifies a complete expressed prokaryotic proteome” ................................................................. 60

3.3. RESEARCH ARTICLE III .......................................... 73

3.3.1. Statement of the own participation ........................... 73

3.3.2. Manuscript: “Dual input control: Activation of Bartonella henselae VirB/VirD4 Type IV secretion system by the stringent sigma factor RpoH1 and the BatR/BatS two component system” ................................................................. 74
3.4. **RESEARCH ARTICLE IV** .................................................................................................................................. 95

3.4.1. Statement of the own participation ........................................................................................................ 95

3.4.2. Manuscript: “Conjugative DNA transfer into human cells by the VirB/VirD4
type IV secretion system of the bacterial pathogen *Bartonella henselae*” ............................... 96

3.5 **UNPUBLISHED RESULTS RELATED TO RESEARCH ARTICLE I** ........................................ 103

3.5.1 Genetic analysis of the *batR-batS* locus ................................................................................................. 103

3.5.2 Initial characterization of BatR transcriptional regulation ................................................................. 106

3.5.3 Differential induction kinetics between P_virB and P_bepD in response to BatR expression ........... 110

3.5.4 Refinement of BatR binding motif at P_virB ......................................................................................... 112

3.6 **UNPUBLISHED RESULTS RELATED TO RESEARCH ARTICLE III** ..................................... 119

3.6.1 Functional characterization of *B. henselae* transposon mutants affected in P_virB
regulation ............................................................................................................................................. 119

3.6.2 LC-MS analysis of *B. henselae* proteins expression in a subset of transposon mutants ............ 120

3.6.3 Rescue of mutant phenotype by BatR or RpoH1 overexpression ..................................................... 124

3.6.4 Results overview and possible roles of the targeted genes in P_virB regulation .......................... 125

3.7 **EXPERIMENTAL PROCEDURES related to unpublished results** ............................................. 130

3.8 **REFERENCES related to unpublished results** ................................................................................. 136

4. **CONCLUDING REMARKS** ............................................................................................................. 141

5. **ACKNOLEDGEMENTS** ..................................................................................................................... 151

6. **CURRICULUM VITAE** ...................................................................................................................... 155
1. Introduction
1.1. BACKGROUND

The adaptive relationship with the environment is a *sine qua non* condition for any intelligent system [1]. Considering the fantastic versatility of bacteria in respect to their lifestyles and possible ecological niches, it is not surprising that these organisms have evolved a vast panel of strategies to sense changes in their environment and to modulate the expression of their genetic repertoire in response to these. The complexity of the lifestyle and the ecological diversity typically positively correlates with the size of bacterial genomes [2,3] and contributes to the complexity of its encoded signal transduction mechanisms ([4] and references therein). For instance, there is a robust correlation between the number of transcription regulators and the square of the total proteins encoded in a bacterial genome [5,6]. This results in a very steep increase in the hierarchical complexity of gene regulation relative to bacterial genome size. Of course, numerous deviations to this general rule have been observed [7], reflecting specific adaptation strategies for given organisms.

In this introduction, we will address the mechanisms of transcriptional regulation by using the bacterial two components systems as a paradigm for signal transduction modules. We will first review the core functionalities of these systems, with their specific mechanisms of activation and modes of action. We will then address some broader aspects such as specificity and modulation of these systems, which are conceptually relevant for most transcription regulators. Finally, we will shortly review two key methods used to gain a genome wide understanding of these regulators and conclude with a perspective on global transcription networks.

1.2. SIGNAL TRANSDUCTION MODULES

Here we will focus on bacterial signal transduction proteins that are involved in adaptation by acting at the level of transcriptional regulation. The basic transcriptional modules that constitute most regulatory circuits can be classified in two majors categories based on their respective architecture: the one component regulatory systems (OCSs) and the two-components regulatory systems (TCSs) [5]. OCSs are by far the most abundant and the most diverse signal transduction modules in bacteria. They typically consist of the combination of an input and an output domain within a single soluble protein molecule [5].
The repressor of the lactose operon (LacI [8]) and the catabolite activator protein (CAP [9]) of *Escherichia coli* constitute prominent examples of such regulators.

The second largest family of bacterial signal transduction modules and the best studied to date is represented by TCSs [4,10]. A prototypical TCS can be divided in four domains, which are distributed on two proteins. The first component – the histidine kinase (HK) – is generally a membrane bound sensor that is composed of an input domain fused to a transmitter domain. The second component – the response regulator (RR) – consists of a receiver domain coupled to an output domain. The RR is localized to the cytoplasm and mediates the adaptive response of the TCS [11,12]. Both components are linked by phosphorylation (see below). Although the origin of TCSs remains a debated topic with uncertain conclusions (reviewed in [13]), several theories have emerged, one of which suggests that these systems have evolved from OCSs. This theory is based on the observation that the vast majority of OCSs containing a DNA-binding motif (that represent the ancestral type of regulatory protein) are predicted to be localized in the cytoplasm [5]. This indicates that OCSs are almost exclusively restricted to detect their stimuli in the cytosol of the organism expressing them and thus restraining perception of environmental cues to light, gases and other small molecules [1]. The emergence of TCSs is proposed to represent an important evolutionary step that overcame the limitations of these single signaling elements by separating the input and the output domains in two individual components, which further evolved into bacterial TCSs [5].

1.3. THE PARADIGM OF BACTERIAL TCSs

With the exception of *Mycoplasma* and few obligate intracellular symbionts where they are absent, most eubacteria encode between few to more than a hundred different TCSs [4,10]. They are also present in some archaea, plants and lower eukaryotes but were lost in metazoans. TCSs have been associated with virtually all processes of bacterial physiology, including various metabolic processes, stress responses, pathogenicity or complex developmental life cycle [4]. They also respond to a very large panel of specific input signals [14]. Phylogenetic analysis showed that TCSs diversity arose by a combination of horizontal gene transfer and lineage-specific expansion combined with domain shuffling, allowing adaptation to new niches [13,15]. Although reasonable understanding of any specific TCS does require individual characterization [16], several
characteristics are shared among this versatile family of signaling molecules, some of which will be summarized here.

As mentioned, a prototypical TCS consists of a membrane bound HK, containing a periplasmic sensory domain and a conserved kinase core that acts as transmitter domain, and a soluble RR, which consist of a conserved N-terminal receiver domain coupled to an C-terminal effector domain ([12] and references therein). The control in two-component pathways is largely dependent on the ability of the HK to regulate the phosphorylation state of its cognate RR. Typically, HKs undergo autophosphorylation on a conserved histidine residue upon perception of an extracellular signal by their periplasmic domain. The phosphoryl group is then transferred to an aspartate residue on the receiver domain of the cognate response regulator. This step triggers a conformational change that usually promotes homodimerization of the receiver domain which in turn modulates the activity of the RR (reviewed in [13,17]). The activated RR then elicits a specific response through its output domain in reply to the input signal.

1.3.1. Domain organization

Bacterial RRs generally share the same domain organization, which consists of a conserved input domain and a variable effector/output domain. The function of the output domain is used for further categorization of RRs. More than 60% of all identified RRs harbor a DNA-binding output domain and are thus classified as transcription factors (TFs). These are further divided in several subfamilies based on the properties of their DNA-binding domain, the three most abundant of which are the OmpR/PhoB, the NalL/FixJ, and the NtrC/DctD subfamilies [16]. The output domains of the remaining RRs include a variety of enzymatic activities, such as diguanylate cyclase and methyl transferase, or protein-protein interaction and RNA-binding effector domains (Fig. 1.1). A noteworthy exception to this canonical organization is constituted by single domain response regulators. These proteins consist of isolated receiver domain and have been associated with numerous regulatory functions [18]. For this introduction, we will consider the domain architecture consisting of a receiver domain and a DNA-binding domain of the OmpR/PhoB subfamily as prototypical for a RR.

In contrast to RRs, bacterial HKs display a highly modular organization. They all share a transmitter domain, which is consist of a dimerization domain, hosting the histidine that is phosphorylated (DHp domain) and a C-terminal catalytic domain (CA) which binds
ATP and harbor the histidine kinase activity [19,20]. Besides this conserved element, HKs are characterized by a wide variety of possible sensory input modules and signal transduction domains [21]. They can also incorporate additional elements, including further phosphotransfer modules, resulting in a complex phosphorelay (Fig. 1.1) [14]. Nearly 25% of the HKs harbor such extended composition and are referred to as hybrid HKs [22].

![Diagram of signal transduction](image)

**Figure 1.1: Domain organization and signaling in TCSs.** A. Prototypic TCS composed of a soluble RR and a membrane bound HK. The RR is composed of a receiver domain (REC) and an effector domain. The HK is composed of a sensor domain, two transmembrane helices (TM1 and TM2) and a cytoplasmic domain, which is subdivided into DHp and CA domains. The path of the phosphoryl group from the ATP to the conserved aspartate (H) within the DHp and final phosphorylation of a conserved aspartate (D) on the RR is highlighted by a red arrow. The frequency of the most common effector domains is indicated. B. Schematic of complex phosphorelay system, with additional domains such as a cytosolic sensor domain (Sc), or a covalently bound REC and a histidine phosphotransfer domain (Hpt) downstream from the CA domain. The blue arrow indicates possible interaction site for modulatory proteins. Adapted from Casino et al., [23].

Determination of the signal(s) recognized by HKs constitutes a very active field of research [14]. There are however large discrepancies in the understanding of individual systems. Only few HKs are understood at atomic resolution, in contrast to many HKs for which even the nature of the activating signal has not yet been unambiguously identified. Mascher *et al.* have proposed a general classification of the HK based on the predicted
localization of their sensing domain (periplasmic, membrane or cytoplasmic) with further sub-categories based on the presence of various signature motifs [21]. However, the exact sensing mechanism remains elusive for most HKs (see below).

1.3.2. General signal transduction mechanism

The activity of a prototypical TCS is determined by the concentration of phosphorylated RR, which is dependent on three phosphotransfer reactions: the autophosphorylation of the HK, the phosphorylation of the RR, and its dephosphorylation [12,21]. The autophosphorylation of the HK generally occurs in trans within a homodimer [14] and is regulated by the input signal. The phosphotransfer on the RR is catalyzed by the RR [12] and the life time of the phosphorylated RR, which can range from seconds to hours, can be regulated by two activities: an intrinsic autophosphatase activity and/or the phosphatase activity of some HKs, although the physiological relevance of the later has been questioned [24]. In some cases, signal perception rather stimulates the phosphatase activity of the HK, as shown for the Cpx TCS of *E. coli* [25].

1.3.3. Activation of HKs upon signal recognition

In order to activate its cognate RR, the HK must first undergo an autophosphorylation reaction, which is triggered by the recognition of the input signal by is sensory domain. The autophosphorylation is one of the three core activities of HKs (autokinase, phosphotransferase and phosphatase) that are believed to be governed by conserved mechanisms that involve the relative positioning of the catalytic active CA domain and the dimeric helical DHp domain, which contains the conserved histidine that becomes autophosphorylated during the activation process [23].

The initial activation steps, *i.e.* the mechanism(s) of signal recognition and transduction by the sensory domain of HKs are globally poorly understood and constitute a source of controversy [14,21]. An interesting example is given by the TCS EnvZ/OmpR of *E. coli*. This extensively studied signal transduction system is responsible for the regulation of the outer membrane porins OmpF and OmpC in response to osmotic stress [26]. EnvZ has a classical HK domain organization with two transmembrane regions and a periplasmic domain that was long believed to be involved in signal perception. However, the mechanism by which EnvZ does respond to changes in osmotic pressure has remained unknown for more than 20 years. A recent study has demonstrated the capacity of the
cytoplasmic domain of EnvZ to sense changes in osmolarity in vitro and in vivo, ruling out the involvement of the periplasmic domain in this process [27]. Further, by using amide hydrogen/deuterium exchange mass spectrometry (HDXMS), the osmosensing module of EnvZ could be assigned to a four-helix bundle within the cytoplasmic dimerization domain of EnvZ. Whether this novel activation mechanism through the destabilization of a cytoplasmic four-helix bundle [27] can be extended to other HKs remains to be demonstrated. This mechanism of activation nevertheless represents an interesting alternative to the piston model (Fig. 1.2) [23]. Based on numerous HK crystal structures [28], this model proposes that changes in the extracellular sensor domain of the HK are transduced to the cytoplasmic domain by a piston-like motion within the connecting helix, which triggers the repositioning of the DHp and the CA domain and allows the switch between the autokinase, the phosphotransferase and the phosphatase activities of the HK [23].

Figure 1.2: Piston model for the regulation of sensory HKs. Signal perception by the sensor domain of the HK results in a rotation or a piston-like motion in the helix that connects this domain to the transmembrane helix 2 (TM2). The signal is further transduced to the helix 1 of the DHp domain (α1), inducing a rotation of the cytoplasmic helix (α2), which in turn results in an alteration of DHp dimer and a displacement of the CA domains. The relative positioning of CA and DHp determines the three functional states of the HK (kinase, phosphotransferase and phosphatase). Adapted from Casino et al., [23].
Interestingly, some HKs have been shown to respond to more than one signal, suggesting that the same protein could be activated by different mechanisms. The best characterized example is certainly the HK PhoQ that together with its cognate RR PhoP constitutes an essential component for the virulence of *Salmonella enterica* [29] and other Enterobacteriaceae [30]. PhoQ displays a prototypical domain organization that comprises a short periplasmic domain and has been shown to integrate at least three different host related signals: low divalent cations concentration [31], sub-lethal concentrations of antimicrobial peptides [32] and acidic pH [33]. NMR measurements indicate that all signals are directly recognized by the periplasmic domain of PhoQ, although through different mechanisms [33]. This example illustrates how a single molecular sensor enables a bacterial pathogen to mount an adaptive response by integration of distinct host signals.

### 1.3.4. Activation of RRs by HKs

The activation of a RR through phosphorylation of a conserved aspartate by its cognate HK represents the next step in the classical TCS signal transduction pathway. It is generally accepted, particularly when referring to RRs involved in transcription regulation, that these proteins exist in an active and an inactive state at equilibrium, and that phosphorylation of their receiver domain mediates and/or stabilizes their active state [34]. However, many different activation mechanisms have been described [12]. Generally, activation of phosphorylated RRs is linked to their homodimerization. Activation thus occurs by phosphorylation-dependent dimerization of the RR, resulting in a DNA-binding competent form. Homodimerization of RRs involves a conserved structural module, the so-called α4-β5-α5 part of the receiver domain, which is especially conserved in the OmpR/PhoB sub-family of RRs [17,35]. For these RRs, numerous NMR and crystallization studies have led to a general phosphorylation dependent activation model, in which a phosphorylated receiver domain forms a two-fold symmetry dimer using the α4-β5-α5 interface. This arrangement in turn favors the formation of the DNA-binding competent head-to-tail arrangement of the two output domains [36,37] (see below).

### 1.3.5. Activation of the OmpR/PhoB subfamily

As previously mentioned, the OmpR/PhoB subfamily represents the most abundant class of DNA-binding RRs (about 30% of all RRs [17]). This subfamily is characterized by the presence of an N-terminal winged helix-turn-helix (wHTH) DNA-binding domain [38]. In contrast to the core activities of HKs and RRs, which share many conserved properties,
a common mechanism for the activation of RR s is still debated. For instance, the *E. coli* RR s OmpR and PhoB were long thought to have fundamentally different modes of action. This impeded the elaboration of a unified model for DNA binding and transcription regulation that would reflect their very similar structural properties [39–41]. Phosphorylation of *E. coli* PhoB is known to induce dimerization of the protein, resulting in a strong increase of the RR affinity for its target DNA [42]. This contributed to the elaboration of the canonical model of RR activation, *i.e.* the phosphorylation dependent DNA-binding of a RR dimer that results in transcription activation. In contrast, dimers of *E. coli* OmpR were never observed in solution, irrespectively to the phosphorylation state of the RR. Moreover, OmpR phosphorylation was described to strongly increase upon DNA-binding *in vitro*. These results led to a model where OmpR dimerization and phosphorylation only occurs upon DNA interaction, and that DNA binding would be a pre-requisite for OmpR activation (reviewed in [34]). A recent study using a combination of sensitive biophysical methods *in vitro* and phosphorylation measurements *in vivo* convincingly demonstrated that OmpR actually follows the canonical mode of activation described for PhoB [34]. The authors propose that previous studies have been compromised by the very low solubility of phosphorylated PhoB, which precipitates at concentrations higher than 10 μM. These data support a conserved mode of activation for OmpR and PhoB that may also extend to the eponymous family of RR s, although differences within individual systems cannot be excluded.

1.3.6. DNA binding by the OmpR/PhoB family of RR s

The DNA-binding properties of the wHTH domain of RR s has been extensively studied and resolved in great details for several individual systems. One of the best characterized systems is the *E. coli* response regulator PhoB, and its interaction with DNA has been resolved at the atomic level [43,44]. The PhoB/PhoR TCS is involved in the regulation of gene expression in response to the external concentration of inorganic phosphate in a board range of bacteria [45,46]. The DNA-binding motif of PhoB – the *pho* box – is an 18-20 base pair (bp) motif largely conserved among proteobacteria. It consists of two direct repeats of 7-11 bp separated by a 4 bp spacer region [45]. PhoB-regulated promoters contain 1-3 *pho* boxes in which the first direct repeat show the highest sequence conservation. Analysis of the crystal structure of the PhoB DNA-binding domain in complex with its target DNA [43] has revealed a head-to-tail arrangement of the monomers binding successive direct-repeat sequences, each repeat corresponding to one turn of the
The sequence variation observed in the sequence of the \textit{pho} boxes suggests that this RR also recognizes the overall shape of the DNA. Integrating the structural information on the receiver domain of PhoB [36], a head-to-head symmetry of the activated receiver domain coupled with a head-to-tail symmetry of the DNA-binding domain is proposed to represent activated form of RRs from the OmpR/PhoB sub-family (Fig. 1.3). The head-to-tail oligomerization of the RR could favor the binding to promoters with multiple boxes [43].

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{Inactive and active domain arrangements in the OmpR/PhoB subfamily of RRs. Although characterized by a different domain orientation in their inactive state, RRs of the OmpR/PhoB subfamily share a common dimeric active state, with a head-to-head orientation for their receiver domain and a head-to-tail orientation for the wHTH domain bound to DNA. The \(\alpha4-\beta5-\alpha5\) regions are highlighted in green and the DNA-binding domains are shown in gold. From Gao \textit{et al.}, [16].}
\end{figure}

Although several members of the OmpR/PhoB subfamily are characterized by a well defined DNA-binding motif (e.g. [29,47]), this is not a general rule. For instance, the sites recognized by \textit{E. coli} OmpR share only very limited homology [48], which can be explained by the few specific contacts that are made between the protein and the DNA [41]. Studies in \textit{Salmonella enterica} have also revealed a strong influence of DNA
supercoiling on the expression of OmpR regulated genes [49]. Together with the observed affinity of OmpR for non-specific DNA, OmpR was proposed to function similarly to nucleoid-associated proteins, with highly degenerated OmpR motifs serving as nucleation points for the cooperative recruitment of further OmpR proteins [49].

1.3.7. Transcription activation by RRs

Although the DNA binding properties of individual RRs have been largely documented, only few studies have convincingly addressed the mechanism of transcriptional activation driven by an activated RR upon binding to its target promoter. It is therefore difficult to come up with a generalized mechanism for this activation. The best characterized mechanism of transcription activation has been resolved for E. coli PhoB [44]. The DNA recognized by this RR - the pho boxes - are usually located 10 bp upstream of the -10 region of the promoter, substituting the -35 sequence in the promoter. As a consequence, the RNA polymerase (RNAP) does not recognize PhoB regulated promoters unless PhoB is bound to them. This led to a model where PhoB promotes the transcription initiation by interacting with the σ^70 subunit of the RNAP [50]. The recent resolution of the crystal structure of a transcription initiation subcomplex that includes the σ_4 domain of the E. coli σ^70 RNAP factor fused with part of the RNAP β subunit, a tandem dimer of the PhoB effector domain and a DNA pho box [44] confirmed this view. Analysis of this structure also revealed that PhoB forms a ternary complex with the σ^70 and the pho box. This association mediates new contacts between the σ_4 domain of σ^70 and the PhoB-bond DNA where a typical -35 sequence would be found. Resolution of this complex further suggests that in addition to the recruitment of the RNAP, PhoB would further act as a transcription activator by facilitating transcript release by remodeling the σ_4 domain of the RNAP [44].

This characterization defines PhoB as a Class II transcription activator (Fig. 1.4, [51]). These factors are characterized by the binding to the -35 region of the target promoter and in most case to activate transcription by making direct contact with the domain 4 of the σ^70 [52]. In contrast, other RRs of the OmpR/PhoB sub-family such as OmpR [53] have been show to act as Class I activators, which are characterized by their interaction with the C-terminal domain of the RNAP α-subunit (αCTD) to activate transcription. These findings illustrate the diversity of transcription activation mechanisms that can be controlled by RRs. A recent study on PhoP mediated transcription activation in S. enterica serovar Typhimurium further highlighted this versatility [54]. In this bacterial
pathogen, not less than 5 promoter architectures are associated with the different PhoP-regulated genes. Each group of promoters is defined by different number of pho boxes, with specific location, orientation, and phasing. Interestingly, the positioning of the PhoP boxes in respect to the -35 and -10 hexamers strongly suggests that PhoP can act as Class I or Class II activator (Fig. 1.4) depending on the promoters [54].

Figure 1.4: Recruitment of the RNAP for different promoter architecture. (a) Simple promoters of Class I and Class II, that depends on the binding of a single activator to recruit the RNAP. The position of the transcription start site (+1) and the -10 and -35 elements are indicated. αCTD: carboxy-terminal domain of the RNAP α subunit, connected to the RNAP by flexible linkers. (b) Complex promoters that require independent contacts by two activators at tandem binding sites to recruit the RNAP. From Barnard et al., [51].

1.4. SPECIFICITY IN TCS SIGNALING

We have addressed the specificity of DNA-binding RRs for their binding sites. The co-occurrence of dozens of RRs and HKs is most bacteria also raises the question of the interaction specificity between HKs and RRs. The classical view of TCSs accounts for a specific association between a HK and its cognate RR, which are typically encoded as an operon. Systematic analysis of the TCSs of Caulobacter crescentus [55] and E. coli [56] have validated this specificity in vitro, with most HKs showing specificity for a single RR substrate. The “one-to-many” (activation of several RRs by a single HKs) and the “many-to-one” (activation of a RR by several HKs) arrangements were also observed albeit to a low frequency [57]. These studies have also demonstrated that the specificity is an inherent property of TCSs and thus sequence encoded rather than a result of protein scaffolding or recruitment of accessory proteins [13]. Based on the analysis of co-variant residues in
cognate HK-RR pairs, the key residues conferring this specificity have been mapped for E. coli TCSs. Direct mutagenesis of these residues allowed rewiring of TCSs specificity in vitro and in vivo [58]. A striking example is provided by the analysis of the two closely related TCSs of E. coli. EnvZ/OmpR and RtsB/RtsA are likely to have evolved by gene duplication followed by sequence diversification. As little as three amino acids substitution within the DHp domain of EnvZ were sufficient to change the substrate specificity of the HK EnvZ to the one of RtsB [58].

These findings raise fundamental questions about the evolution of TCSs, for instance how newly duplicated TCS can diversify to become isolated from the original system while maintaining high phosphotransfer efficiency. This question was addressed by conducting a systematic analysis of the sequence space defining the signaling specificity between the E. coli TCSs EnvZ/OmpR and RtsB/RtsA (3 residues on the HKs and 6 residues on the RRs) [59]. This study demonstrated that it is possible to replace the residues conferring the specificity to the EnvZ/OmpR TCS by the ones of the RtsB/RtsA by a sequence of ordered mutations such that the cognate proteins would maintain a high level of specific phosphotransfer. Further, this transition did not cross the sequence space occupied by a third related TCS, CpxA/CpxR. Besides fascinating insight into the evolution and diversification of TCS this trajectory-scanning mutagenesis also opens interesting application for the targeted rewiring of TCSs.

1.5. MODULATION OF TCSs BY ACCESSORY PROTEINS

In the recent years, many reports describing the involvement of accessory proteins in the modulation of TCS signal transduction have been published, giving raise to the emerging concept of three-component systems. HKs and RRs are easily identifiable from genomic sequences by their conserved signature motifs. In contrast, accessory proteins involved in TCS signaling are very difficult to predict as they belong to very diverse families and often represent specific adaptations for a given organism [60]. Accessory proteins where shown to influence the activity of TCSs at many different level of the signal transduction cascade(depicted by arrows in Fig. 1.1). Their mode of action ranges from the modulation of signal perception by HKs or their subsequent activation to the regulation of the RR phosphorylation state [14,61]. Here we will only discuss two distinct examples of such modulatory proteins - the small soluble protein PmrD and the periplasmic proteins ExoR.
In *S. enterica*, resistance to polycationic antimicrobial peptides is primarily controlled by the PmrA/B TCS, which regulates genes involved in the control of lipopolysaccharide modification [62] in response to extracellular Fe$^{3+}$ concentrations [63]. The PmrA regulated genes are however also activated under the conditions that activate PhoP/PhoQ, another TCS that responds to low extracellular Mg$^{2+}$ concentrations, antimicrobial peptides and acidic pH [64]. This cross-regulation is mediated by an 85 amino acids protein called PmrD [65] that is expressed under the control of the PhoP/PhoQ TCS system. PmrD binds phosphorylated PmrA and prevents its dephosphorylation by its cognate HK PmrB [66]. As this interaction does not interfere with downstream signaling, it promotes PmrA dependent transcription (Fig. 1.5). Phosphorylated PmrA also negatively regulates the expression of PmrD by binding to its promoter, which prevents excessive levels of activated PmrA [67]. Such regulators, also referred to as TCS connectors [61], enable the establishment of regulatory links between independent signal transduction pathways, thus allowing a better adaptive response to changes in the environment.

**Figure 1.5: Role of PmrD as TCS connector in *S. enterica*.** The model illustrates how PmrD allows the integration of the PhoP/PhoQ signaling for the expression of the PmrA/PmrB regulated gene *pbgP*. See main text for detailed explanation. From Kato et Groisman [66].

Another example of TCS modulation by an accessory protein is provided by ExoR and the ExoS/ChvI TCS. This TCS is essential for the α-proteobacterium *Sinorhizobium meliloti* to establish nitrogen fixing symbiosis with its plant host by regulating the production of exopolysaccharides [68]. In this organism, ExoR, a small periplasmic protein was shown to interact with the periplasmic domain of the HK ExoS and to inhibit its activity. Consequently, deletion of *exoR* triggers overexpression of exopolysaccharides, as
a result of the hyper-activation of the ExoS/ChvI TCS [69,70]. ExoR interaction with the sensory domain of ExoS maintains the HK in a signaling off-state, preventing downstream signaling. Interestingly, exoR transcription is activated by ChvI whereas the relief of ExoS inhibition is controlled at a post-translational level, presumably through the proteolysis of ExoR in response to a yet elusive host signal by periplasmic proteases [71]. This regulation mechanism is supported by similar finding in the plant pathogen Agrobacterium tumefaciens. In this organism, the orthologous TCS ChvG/ChvI is also repressed by ExoR [72] and activation of the HK ChvG correlates with the proteolysis of ExoR protein, both taking place at acidic pH [72,73]. This indicates that ExoR may play an active role in the signal recognition mechanism of this family of TCSs.

1.6. GENOME-WIDE APPROACHES TO STUDY TCSs

The developments of DNA sequencing technologies within the last decade [74] and the resulting explosion of available sequenced microbial genomes have profoundly impacted the field of transcriptional regulation. Besides the emergence of new research areas such a comparative genomics, phylogenomics and numerous other “omics”, this technologic revolution has also provided a genome wide perspective for most bacterial models. The continuous development of bioinformatics and sequence analysis tools and databases (e.g. [75–77]) has enabled the automated prediction of the entire repertoire of TFs encoded within any sequenced species [10,78]. However, despite continuous progress in predictive sciences, empirical experimentation is still required to identify the transcriptional targets of transcriptional regulators, and to further connect these into transcriptional network(s). Here we will first shortly discuss two generic approaches that allow addressing this type of questions at the genome level: genome wide transcription analysis and genome wide mapping of TF binding sites. We will shortly describe the underlying technologies and provide some examples illustrating how these approaches have contributed to understand TCS signaling. We will then conclude by showing how such genome wide approaches enable to study the evolution of TCS driven transcriptional circuits.
1.6.1. Genome wide transcription analysis applied to TCS signaling

The availability of full genome sequence has raised a growing interest for the genome wide changes in transcription profiles of the encoded genes in response to different stimuli, and the identification of the underlying transcription regulator(s). Determination at the genome scale of the co-regulated genes in response to specific stimuli (stimulon) and identification of genes under the control of a given regulator (regulon) constitute by now classical aspects of transcription regulation research. A breakthrough to obtain such global information was made with the establishment of DNA-microarrays. This technology relies on the hybridization of a labeled nucleic acid sample to a set of DNA probes immobilized on a solid substrate (reviewed in [79–81]). A prototypical DNA microarray consists of single stranded DNA molecules of known sequences immobilized at a defined location on a solid substrate. The latest technological developments using Digital Micromirror Devices together with parallel combinatorial DNA synthesis chemistry allow the synthesis of 385’000 to 4.2 million unique probe features in a single array [82]. Such arrays have recently been used for massive parallel transcription profiling of *Bacillus subtilis* exposed to 104 different conditions, providing unique insights into the regulatory architecture of this organism [83].

Transcription profiling using DNA microarrays requires the reverse transcription of the extracted mRNA and the incorporation of a label in the resulting cDNA for subsequent detection of the hybridized product. Although widely used in the last decade, this technology suffers from a major limitation, since it relies on the *a priori* knowledge of the DNA sequence of the studies organism. This constrain is especially limiting for research on prokaryotes, as these organisms can display remarkable genomic plasticity between different isolates or even within a clonal population [84,85]. The recent development of RNA-seq, an approach that is derived from the next generation sequencing technology, allowed overcoming this limitation [86]. As RNA-seq consists in the *de novo* deep-sequencing of the cDNA sample matching the RNA of interest, it allows the identification and quantification of any transcripts present in a sample, including small regulatory RNAs, which abundance in bacteria had long been underestimated [87,88]. Further, this methodology does not suffer from the limitations of competitive hybridization encountered with DNA microarrays [86]. Moreover, specific developments of this technique allow the genome wide mapping of transcriptional start sites at the nucleotide level [89].
Comparing the transcription profiles of a mutant to its parental strain constitutes a generic approach to study the function of any TCS. Such experiments ideally allow the determination of the **regulon** for this given TCS, *i.e.* the sum of genes whose expression is controlled by that system, either directly or indirectly. An alternative setup is to compare a wild-type strain with a derivative that over-expresses a constitutive active allele of the response regulator. This approach proves particularly useful to study TCSs that control essential functions and thus cannot be deleted. A large number of studies report the use of transcription profiling to study the function of TCSs, and the regulon of many RRs has been resolved in great details in numerous bacteria (*e.g.* [29,90,91]). Yet only few studies have addressed this question in a systematic manner. This was performed for *E. coli*, where the transcription profiles of mutants in the 36 TCSs it encodes have been compared [92]. This analysis revealed the different extent of perturbation caused by the deletion of the different TCSs. In more than half of the cases, the deletion only affected a small number of genes, while only few TCSs behaved as global regulators. Furthermore, it revealed cross-regulations and cascade regulations between different TCS regulons and highlighted that the regulation of central cellular processes *e.g.*, flagellar synthesis, control of the RpoS regulon or maltose transport, integrates inputs from multiple TCSs. Such experiments have contributed to the understanding of the contribution of TCS signaling to the global transcriptional regulatory networks (see also 1.7), with hierarchies of transcriptional regulators and critical nods for signal integration [93].

### 1.6.2. Genome wide mapping of RRs DNA binding sites

Genome wide transcription analysis of regulatory mutants provides a global knowledge of the processes these factors control and can help to understand the signal(s) they respond to. However, this approach doesn’t allow the discrimination between direct and indirect effects. Indeed, TFs – including TCSs – are commonly organized in transcription networks. Deletion of one factor may have very indirect effects on the global transcription profile as a consequence of its requirement for the expression of other regulators. The determination of the DNA sequence(s) recognized and bound by transcriptional regulators represents a complementary approach to characterize these factors. The classical approach consists in assessing the ability of a purified regulator to bind to a labeled DNA probe carrying the putative DNA-binding site *in vitro*. This is the principle of the electrophoretic mobility shift assay (EMSA) [94]. This technique depends
on the *a priori* knowledge of the genes regulated by the factor of interest, which can be approached by transcriptional profiling. This method can be further combined with DNA footprinting [95], a technique that relies on the ability of the regulator to protect the DNA it binds from DNA degrading condition *e.g.*, DNase I digestion. This approach allows the determination of the nucleotides bound by a TF and can constitute the basis for the genome wide determination of DNA-binding sites when combined with *in silico* predictions and further EMSA iterations [96].

An alternative strategy is based on the selective immunoprecipitation of the regulator bound to its target DNA, also referred to as chromatin immunoprecipitation (ChIP). This technique offers the great advantage that it can be performed *in vivo* and allows a time resolution of the binding. Therefore it doesn’t only allow determining *where* a regulator is binding, it also allow addressing *when* or in response to which treatment/stimulus this binding is taking place. Identification of the DNA sequences recovered by immunoprecipitation was first greatly facilitated by the use of the microarray technology (ChIP-on-chip) and more recently by the next generation sequencing (ChIP-seq). The ChIP-on-chip approach relies on the hybridization of the recovered DNA on a microarray after fluorescent labeling [97] whereas the ChIP-seq simply consist of the deep sequencing of the recovered DNA and subsequent mapping to the genomic sequence [98]. These techniques are undergoing very fast development and new applications are published on a regular basis. For instance, the combination of ChIP-seq with the use of DNA footprinting with an exonuclease (ChIP-exo) allows the narrow resolution of the DNA-binding sites (Fig. 1.6, [99]).
Figure 1.6: Single base-pair resolution of ChIP-exo. (A) Illustration of the ChIP-exo approach, with treatment of the bound DNA from a ChIP with a 5'-3' exonuclease. (B) Comparison of ChIP-exo, ChIP-chip and ChIP-seq for the yeast regulator Reb1 at specific loci and (C) aggregated Reb1 signals for 791 binding sites. Adapted from Rhee et al. [99]

EMSA based approaches have been used for the characterization of numerous DNA-binding RRs in various organisms, and have contributed to the characterization of many TCSs (e.g. [63,100,101]). The combination of ChIP and transcription profiling has eased the genome wide characterization of TFs binding sites. This approach has already allowed the determination or the refinement of several regulons, such as the SsrB regulon in S. enterica [102], the OmpR regulon in S. typhi [103] or the PhoP regulon in E. coli, S. enterica [104], and in Yersinia pestis [30]. This type of analysis further enables to study the evolution of the transcription networks between related bacteria.

A revealing comparison performed on the PhoP regulon in S. enterica (22 transcription units) and Y. pestis (18 transcription units) revealed that only 3 transcription units are directly controlled by this RR in both species [30]. The main reason for this divergence is that most of the PhoP regulated genes in Y. pestis or in S. enterica have no homolog in the other species, or even in other Enterobacteriaceae [30]. The only three genes controlled by PhoP in both species are those encoding the most conserved proteins of respective regulon in either species. These are the RR PhoP, its cognate HK PhoQ, and SlyB, a lipoprotein that negatively regulates the activity of the PhoP/PhoQ TCS. On the other hand, the genes present in both Y. pestis and S. enterica but regulated by PhoP in
only one of the species are indicative of transcriptional rewiring, i.e. change of interaction between orthologous regulators and target genes [30]. The prevalent concept of variable and conserved subsets in the regulons of orthologous TFs is derived from this type of studies. According to this concept, each regulon consists of a set of highly variable species-specific genes, which allow the bacteria to adapt to its environment, and a restricted set of conserved genes that encode the proteins controlling the amount of active TF [105].

1.7. TCSs AND TRANSCRIPTION NETWORKS

TCSs are not autonomous features within a cell. In contrast, they are always part of larger integrated regulatory modules and are intimately embedded within the transcription network of the cell. The study of regulatory interactions between TFs and the genes they regulate can thus be approached from the perspective of network science, an interdisciplinary research field at the inter-phase of mathematics, physics and computer science [106]. Here we will present some basic principles of transcriptional networks organization and some key findings derived from the application of network theories to global transcription control in bacteria.

Complex transcriptional networks are the sum of smaller regulatory entities, or network motifs, composed of basic units (reviewed in [107,108]), which are responsible for the processing of specific information or signals. Basic units can be divided in three different types depending on their mode of regulation (Fig. 1.7A). Simple regulation describes the activation of a TF by a signal, resulting in the direct activation of its target gene. Negative auto-regulation describes TFs that negatively regulate their own transcription whereas positive auto-regulation applies to TFs that positively regulate their own transcription. Prototypical TCSs represent a specific form of positive auto-regulation, termed positive feedback loop (see 1.7.1). Each of these modules is characterized by an intrinsically different signal-response curve, depending on the strength of their promoters and the intensity of the signal [108,109]. These basic units are further organized in higher complex entities, referred to as network motifs or local networks, which are able to perform basic signal integration operations. The three prevalent network motifs in transcription regulation are single input (SIM), multiple input (MIM), and feed-forward loop (FFL) motifs (Fig. 1.7B).
Figure 1.7: Basic units, network motifs and global regulatory network in bacterial transcription. (A) The three basic units (simple, positive and negative autoregulation) are schematically represented. (B) Schematic representation of the three representative network motifs: the single input (SIM), the multiple inputs (MIM) and the feed forwards loop (FFL) motifs. For the FFL, the two most common organizations are depicted. (C) Global network: transcription regulatory network of Mycobacterium tuberculosis with input nodes shown in blue and intermediate (or transit) nodes shown in green. Adapted from Alon [107] and Babu et al., [108] (A, B), and Balázsi et al., (C, [110]).

The main function of the SIM motif is to ensure the coordinated expression of a set of genes with shared function. Despite its simple arrangement, this motif can display various nuances, with for instance target genes showing different affinities for the TF. Such architecture results in a temporally distinct expression of the regulated genes [107]. Rather, the MIM motif allows the integration of distinct signals and enables combinatorial activation of the regulated genes [109]. The FFL motif consists of three genes: a top-level TF that regulates a target gene and an intermediate regulator, which also regulates the same target gene [107]. FFMs are typically involved in the response to persistent signal and are
characterized by their ability to filter noise. The two most common FFL motifs are depicted in Fig. 1.7B. In a so-called coherent FFL motif, both top-level and intermediate TFs act as positive regulators, whereas in the case of the incoherent FFL motif, the intermediate regulator acts as a repressor. Each of these network organizations are characterized by distinct dynamics in respect to time and signal response [107].

1.7.1. Positive feedback regulation of TCSs

The most common genetic organization for TCSs is to be encoded as a single operon structure under the control of at least two promoters. A constitutive promoter ensures the expression of a basal level of the TCS, which is required to perceive the initial signal, and a second promoter that is positively regulated by the RR [111,112]. This architecture shows all characteristics of a positive feedback loop, where activation of the TCS results in an increased synthesis of its constituents (Fig. 1.8). The specific activation dynamics of any given TCS is determined by the relative strength of both constitutive and activated promoters, the rates of phosphorylation and dephosphorylation of the RR as well its degradation rate and the cooperativity of the RR binding to its own promoter determine [113].

![Diagram of TCS regulation](image_url)

**Figure 1.8: Positive feedback and activation of TCSs.** (A) Schematic representation of a prototypic TCS (R/H) encoded in an operon under the control of a constitutive and an activated promoter (P₁ and P₂). The phosphorylation module determines the concentration of activated RR ([R-P]) as a function of the total sensor (H) and regulator (R) concentration and the activating signal intensity. (B). Influence of the strength of the constitutive promoter on the output levels of the TCS ([R-P]) depicted in (A), in the presence (solid line) or absence (dashed line) of positive feedback. Blue and red lines correspond to low and high stimulus, respectively. Adapted from Mitrophanov et al. [114].
Introduction

Computational analysis has revealed that the strongest effect of positive feedback regulation on downstream signaling mediated by TCSs is reached when the strength of the constitutive promoter is weak [114]. Under these conditions, the positive feedback is predicted to strongly affect the output of a TCS at both high and low stimulus conditions (Fig. 1.8). In contrast, with a strong constitutive promoter, this effect is minimized and may only contribute to the response to elevated stimulus conditions [114]. These predictions were verified using *S. enterica* strains with mutation within the promoter of the *phoP/phoQ* operon [114]. In this organism, positive feedback regulation of the PhoP/PhoQ TCS is also associated with an initial surge (or overshooting) of the PhoP regulated genes in response to stimulation [115]. This fast and intense transcriptional response was shown to be dependent on the positive feedback loop as it was abrogated when the *pho* box of the inducible promoter of the *phoP/phoQ* operon was mutated [115]. The phosphatase activity of PhoQ was also recently shown to contribute to and to be required for this response curve [116]. The strong attenuation observed in the mouse infection model when using a strain of *S. enterica* strain carrying a mutation in the *pho* box of the *phoP/phoQ* operon [115] highlights the importance of this surge for *Salmonella* virulence *in vivo*.

Another feature derived from the positive feedback regulation of TCSs is the ability for the bacteria to develop a memory for the stimulus they respond to. This allows an enhanced sensitivity to subsequent exposure to the same signal. This mechanism was described for the PhoB/PhoR TCS of *E. coli*. The increased levels of HK and RR produced in response to a first exposure to inorganic phosphate results in a higher sensitivity for this molecule in a subsequent exposure [117].

1.7.2. Bistability in the context of TCS regulation

The common outcome from a sensory regulatory network is the mounting of a graded response, with a continuous relationship between the intensity of the input signal and the strength of the response. This type of response is defined as monostable. In contrast, some regulatory networks display bistability, meaning that for a given range of input signal the system can adopt two distinct steady states. Bistable regulators are important components to establish population heterogeneity [112]. The minimal requirement for a regulatory circuit to display bistability is the presence of a positive feedback loop [113]. Considering that most TCS are controlled by a positive feedback, it raises the question whether these signaling modules can sustain a bistable behavior. Mathematical modeling using the well characterized parameters of the *E. coli* EnvZ/OmpR
TCS [118] highlighted some fundamental requirements for TCSs systems to display bistability. According to this model a classical TCS, in which the HK displays a bifunctional kinase/phosphatase activity, can only show a bistable behavior if the unphosphorylated RR and HK form a stable *dead-end* complex and if an independent phosphatase acts upon the phosphorylated RR (Fig. 1.9) [118]. As most of the canonical TCSs described to date do not fulfill these properties, their signaling is predicted to be strictly monostable. Two component signaling is thus rather characterized by a graded response to the perceived stimulus, although bistable behavior can arise from non-prototypical TCS architecture, as it is the case for *B. subtilis* DegU/DegS TCS [119]. This TCS displays a non canonical chromosomal organization, with the RR encoded downstream of the HK and controlled by a separated, autoregulated promoter [120].

---

**Figure 1.9: Bistability in TCSs.** (A) Scheme of a prototypical TCS, with a positive feedback. The concentration of phosphorylated RR dimers (RRP)<sub>2</sub> controls the TCS operon and downstream regulated targets (ds). (B) Schematic representation of the post-transcriptional reactions, where SHKP denotes the phosphorylated HK and RRP the phosphorylated RR. A doted box indicates a possible variation with formation of a *dead-end* complex SHK.RR. (C) Sigmoidal signal-response curve of a monostable system, characterized by a continuous signal-response relationship. (D) Bistability in a bifunctional TCS that comprises an alternative phosphatase and the formation of a dead-end complex. Two steady states coexist between the signal intensity S<sub>1</sub> and S<sub>2</sub>. Adapted from Ray and Igoshin (A, B [121]) and Mitrophanov and Groisman (C, D [114]).
In general signaling terms, a monostable behavior represents the ideal design for most of the TCSs that work as environmental sensors, as it allows the mounting of a graded and reversible response to fluctuations in the level of stimulus. In contrast, the all-or-nothing response that is generally associated with bistable response is well suited for TCSs involved in the regulation of complex directional processes such as cell cycle progression, differentiation, stress responses or specific stages of host-adaptation in the case of bacterial pathogens [119].

1.7.3. Global transcription networks and TCSs

The accumulation of genome wide data that describe the global action of TFs together with the availability of detailed information on individual transcriptional regulators has enabled modelers to reconstruct the global transcription network in several model bacteria such as *B. subtilis* [122], *M. tuberculosis* [110] or *E. coli* [123]. These approaches have revealed some of the interesting properties that govern the higher hierarchy of these networks. For instance, the global transcription network of *E. coli* was reconstructed based on 3000 regulatory interactions comprising 7 sigma factors and 115 TFs [123]. This corresponds to one third of the total regulators encoded in this bacterium. Analysis of this global networks organization reveals a hierarchical organization with non-uniform distribution of the connectivity (reviewed in [109]). Only few TFs act as global regulators, influencing the transcription of multiple genes whereas most TFs only regulate few targets. This so-called scale-free organization confers very high robustness to the network [108]. Interestingly, TCSs, which commonly respond to external signal(s), appear to be more prominent at the bottom of the different regulatory cascades. They are rarely recruited as global regulators, which are more frequently associated with the sensing of endogenous signals such as the metabolic state of the bacteria. At least in *E. coli*, the external input information relayed by TCSs often defines the last regulatory steps of developmental processes, as in the case of motility or biofilm formation [123].
1.8. REFERENCES


2. Aim of the thesis
AIM OF THE THESIS

The primary aim of this thesis, started in July 2009, was the investigation of the molecular mechanisms underlying the adaptive response of the bacterial pathogen *Bartonella henselae* to its host, with a main focus on the regulation of the VirB/D4 Type IV secretion system (T4SS) and its secreted effectors during endothelial cell infection.

To this end, I combined genome wide approaches such as transcriptome and proteome profiling, performed genetic screening and used a panel of more targeted approaches, including the generation of regulatory mutants, the establishment of host-free conditions for the induction of the VirB/D4 T4SS and the monitoring of promoter activity by flow-cytometry.
3. Results
3.1. RESEARCH ARTICLE I (published)

The BatR/BatS Two-Component Regulatory System Controls the Adaptive Response of *Bartonella henselae* during Human Endothelial Cell Infection

Maxime Quebatte, Michaela Dehio, David Tropel, Andrea Basler, Isabella Toller, Guenter Raddatz, Philipp Engel, Hillevi L. Lindroos, Siv G. E. Andersson, and Christoph Dehio


3.1.1. Statement of the own participation

My contribution to the *research article I* included the participation to the design of the microarrays used in this publication and the entire production of these arrays. I performed the experiments to generate the microarray data together with Andrea Basler. I participated to and finalized the analysis of the data including submission to the EMBL database. I planned and performed all experiments presented in Fig. 3, 4, 8 and 9 and Fig. S2, S3 and performed the experiments presented in Fig. 6C and 7C based on preliminary results obtained by David Tropel. I wrote the manuscript together with Siv Anderson and Christoph Dehio and assembled all the figures and tables presented in the manuscript.
The BatR/BatS Two-Component Regulatory System Controls the Adaptive Response of *Bartonella henselae* during Human Endothelial Cell Infection\(^\dagger\dagger\dagger\)

Maxime Quebatte, ¹ Michaela Dehio, ¹ David Tropel, ¹ Andrea Basler, ¹ Isabella Toller, ¹ Guenter Raddatz, ¹ Philipp Engel, ¹ Sonja Huser, ¹ Hermine Schein, ¹ Hillevi L. Lindroos, ² Siv G. E. Andersson, ² and Christoph Dehio¹*

Focal Area Infection Biology, Biozentrum, University of Basel, Klingelbergstrasse 70, 4056 Basel, Switzerland, and Department of Molecular Evolution, Evolutionary Biology Center, Uppsala University, 75236 Uppsala, Sweden²

Received 23 December 2009/Accepted 12 March 2010

Here, we report the first comprehensive study of *Bartonella henselae* gene expression during infection of human endothelial cells. Expression of the main cluster of upregulated genes, comprising the VirB type IV secretion system and its secreted protein substrates, is shown to be under the positive control of the transcriptional regulator BatR. We demonstrate binding of BatR to the promoters of the virB operon and a substrate-encoding gene and provide biochemical evidence that BatR and BatS constitute a functional two-component regulatory system. Moreover, in contrast to the acid-inducible (pH 5.5) homologs ChvG/ChvI of *Agrobacterium tumefaciens*, BatR/BatS are optimally activated at the physiological pH of blood (pH 7.4). By conservation analysis of the BatR regulon, we show that BatR/BatS are uniquely adapted to upregulate a genus-specific virulence regulon during hemotropic infection in mammals. Thus, we propose that BatR/BatS two-component system homologs represent vertically inherited pH sensors that control the expression of horizontally transmitted gene sets critical for the diverse host-associated life styles of the alphaproteobacteria.

The alphaproteobacterium *Bartonella henselae* is a globally distributed zoonotic pathogen that naturally infects cats by causing an asymptomatic intraerythrocytic infection. Transmission to humans can result in various clinical manifestations, including cat scratch disease in immunocompetent patients and bacillary angiomatosis-peliosis as distinct forms of vascular tumor formation characteristically seen in immunocompromised patients (14). The cat flea (*Ctenocephalides felis*) is responsible for cat-to-cat transmission (12), whereas transmission to humans is caused by cat scratches and bites. Notably, *B. henselae* and the closely related human-pathogenic species *Bartonella quintana* and *Bartonella bacilliformis* are unique in the bacterial kingdom for their capacity to cause proliferation of the human vasculature (16). Thus, cultured human endothelial cells (HEC) represent a valid *in vitro* model to study the unique interactions of *B. henselae* with the human vasculature that culminate in the formation of vascular tumors (17, 19).

Type IV secretion systems (T4SSs) are multicomponent transporters crucial for the pathogenesis of many Gram-negative bacteria (e.g., *Helicobacter*, *Legionella*, *Bordetella*, *Brucella*, *Agrobacterium*, and *Bartonella*). Bacteria use these systems to deliver bacterial effector proteins or DNA-protein complexes into the cytoplasm of their host cells in order to subvert their cellular function (13). For *B. henselae*, the VirB T4SS and the *Bartonella* effector proteins (Beps) mediate most of the cellular phenotypes associated with *B. henselae* infection of HEC (46). The translocation of the Beps into HEC mediates (i) a massive rearrangement of the actin cytoskeleton, resulting in the formation of an invasome that internalizes large bacterial aggregates; (ii) nuclear factor kappa B-dependent proinflammatory activation; and (iii) the inhibition of apoptotic cell death (46, 49).

The importance of the VirB T4SS for *Bartonella* pathogenicity was also demonstrated by a large-scale signature-tagged mutagenesis screen in *Bartonella tribocorum* to identify essential genes for the colonization of its rat reservoir host (42). In the same screen, the *batr* gene encoding the putative response regulator of the predicted BatR/BatS two-component system (TCS) was also identified and thus is likely to be involved in the transcriptional regulation of *Bartonella* pathogenicity. Bacterial two-component regulatory systems are a key element of the transcriptional regulatory circuits that enable organisms to elicit an adaptive response to changes in their host-associated microenvironments and to mount the appropriate response to successfully establish mutualistic or pathogenic interactions with their respective hosts (6, 36, 37, 51). The closest homologs of BatR/BatS in the alphaproteobacteria were shown to be essential for effective host interaction. For the facultative intracellular pathogen *Brucella abortus*, the BvrS/BvrR TCS is essential for virulence and is responsible for extensive cell envelope modulation, including the upregulation of the outer membrane proteins Omp3a/Omp3b (25, 32). The ChvG/ChvI TCS of the plant pathogen *Agrobacterium tumefaciens* is essential for plant tumor induction (8, 33) and controls the expression of acid-inducible genes involved in virulence (29, 31). Similarly, the ExoS/ChvI TCS of the legume-nodulating sym-
biont Sinorhizobium meliloti plays an essential role in the establishment of symbioses with its host by regulating the production of succinoglycan (11) and was also shown to regulate transcription of a constitutive active form of ChvI (9).

To date, our knowledge of the coordinated response orchestrated by B. henselae to successfully invade and colonize HEC is sparse, as is our understanding of the environmental signals perceived by the bacteria during host cell interaction. Additionally, the factors implicated in the regulation of the B. henselae adaptive response to host cell interaction remain elusive. To address these important questions, we have used HEC as a model to examine the expression profile of B. henselae during host cell infection, with a specific focus on the BatR/BatS TCS. The data obtained from the characterization of a batR deletion mutant reveal that BatR is essential for B. henselae pathogenicity. By transcriptional-profiling analysis, we demonstrate that batR is required for the upregulation of a critical cluster of genes regulated during the infection of HEC, including genes encoding the VirB T4SS and its cognate secreted effectors (Beps). Evidence is provided that BatR/BatS constitute a functional TCS, and furthermore, that BatR binds directly to the promoters of the virB operon and the bepD gene. BatR is thus the first regulatory protein shown to be directly involved in the regulation of these key pathogenicity factors in Bartonella. Moreover, we show that the BatR/BatS TCS is activated in a neutral pH range (pH 7.0 to 7.8) with an optimum at the physiological pH of blood (pH 7.4). Finally, we show that despite the evolutionary conservation of both the histidine kinase sensor protein and the response regulator across alphaproteobacterial species, a large subset of the BatR/BatS regulon is specific to the genus Bartonella and/or has recently been horizontally transferred into the genus.

**RESULTS**

### MATERIALS AND METHODS

**Bacterial strains, cell lines, and growth conditions.** B. henselae and Escherichia coli strains were grown as previously described (49). Plasmids were introduced into B. henselae by conjugation from E. coli using three-parental mating. Table 1 lists all strains and plasmids used in this study. Table S4 in the supplemental material lists all oligonucleotides used in the study. The endothelial cell line Ea.hy926, resulting from a fusion of human umbilical vein endothelial cells (HUVEC) and the lung carcinoma cell line A549 (20), was cultured as reported previously (27).

**Construction of an in-frame deletion and complementation of the deletion mutant.** The batR in-frame deletion mutant of RSE247 was generated by a two-step gene replacement procedure, as described previously (46, 48), using the suicide plasmid pAB001. The 5' 393-bp and 3' 706-bp flanking regions were amplified with primers prAB007, prAB008, prAB009, and prAB010 and combined by megaprime PCR. The resulting fragment was digested by XbaI and inserted into the site of pTR1000, yielding pAB001. The ΔbatR strain contains an in-frame deletion of 678 bp in batR, resulting in a 45-bp cryptic open reading frame composed of 5' and 3' sequences of batR. To generate the batR complementation plasmid pbatR, a 782-bp fragment containing batR and the Shine-Dalgarno sequence of pPG110 was amplified using primers prT1009

---

**TABLE 1. Bacterial strains and plasmids**

<table>
<thead>
<tr>
<th>Strain or plasmid</th>
<th>Relevant characteristics</th>
<th>Reference or source</th>
</tr>
</thead>
<tbody>
<tr>
<td>E. coli</td>
<td>thrB1004 pro thi strA hsdS lacZΔM15 (F lacZΔM15 lacIΔ proA+ B+ ΔdapA:cem (Erm') pir ) traD636 proA+ ΔB+ ΔgfpΔA hsdR17(k(-) rpsL (-) mcrA-17) rpsL (Str') endA1 gryA406 relA1 lacF Δ lacB' lacI' (Tet')</td>
<td>Invitrogen</td>
</tr>
<tr>
<td>DTHa</td>
<td>AΔ80lacZΔM15 ΔlacZΔM15 ΔlacZYA-argF)169 recA1 endA1 hsdR17(k(-) mcrA-17) supE44 thi-1 gyrA relA1 endA1 hsdR17(k(-) mcrA-17) supE44 thi-1 recA1 gryA406 relA1 lacF Δ lacB' lacI'</td>
<td>Novagen</td>
</tr>
<tr>
<td>NovaBlue</td>
<td>expression construct for His6-BatR (pET15b)</td>
<td>This study</td>
</tr>
<tr>
<td>TOP10</td>
<td>expression construct for His6-BatS expression</td>
<td>This study</td>
</tr>
<tr>
<td>Rosetta(DE3)</td>
<td>expression construct for His6-BatS expression</td>
<td>This study</td>
</tr>
<tr>
<td>B. henselae</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RSE247</td>
<td>Spontaneous Sm' strain of ATCC 49882, serving as wild type</td>
<td>GE Healthcare</td>
</tr>
<tr>
<td>ABB127</td>
<td>batR in-frame deletion mutant of RSE247</td>
<td>This study</td>
</tr>
</tbody>
</table>

**Plasmids**

<table>
<thead>
<tr>
<th>Plasmids</th>
<th>Relevant characteristics</th>
<th>Reference or source</th>
</tr>
</thead>
<tbody>
<tr>
<td>pCR-blunt II TOPO</td>
<td>Cloning vector</td>
<td>Invitrogen</td>
</tr>
<tr>
<td>pTR1000</td>
<td>oriT, oriC, gfp mut2, lacIΔ rpsL. Km'; mutagenesis vector</td>
<td>Novagen</td>
</tr>
<tr>
<td>pAB001</td>
<td>Derivative of pTR1000 used for the generation of a batR in-frame deletion mutant</td>
<td>This study</td>
</tr>
<tr>
<td>pCD341</td>
<td>Km' Tra' mob' oriRIF1010 (IncQ) pUC19, lacIq</td>
<td>This study</td>
</tr>
<tr>
<td>pBatR</td>
<td>batR in pCD341</td>
<td>This study</td>
</tr>
<tr>
<td>pET15b (+)</td>
<td>Cloning and expression vector; N-terminal His tag</td>
<td>Novagen</td>
</tr>
<tr>
<td>pDT009</td>
<td>Expression construct for HisBAT-BatR (pET15b)</td>
<td>This study</td>
</tr>
<tr>
<td>pPG124</td>
<td>Expression construct for HisBepD expression</td>
<td>This study</td>
</tr>
<tr>
<td>pSH007</td>
<td>Expression construct for HisVirB5</td>
<td>This study</td>
</tr>
<tr>
<td>pGEX-2T</td>
<td>Cloning and expression vector; N-terminal GST tag</td>
<td>GE Healthcare</td>
</tr>
<tr>
<td>pDT020</td>
<td>Expression construct for GST-BatSΔ1-332</td>
<td>This study</td>
</tr>
<tr>
<td>pCD366</td>
<td>Promoter probe plasmid</td>
<td>This study</td>
</tr>
<tr>
<td>pIT009</td>
<td>pVirB-gfp, virB promoter region (bp -366 to +21) in pCD366</td>
<td>This study</td>
</tr>
<tr>
<td>pIT011</td>
<td>pVirB-gfp, virB promoter region (bp -333 to +13) in pCD366</td>
<td>This study</td>
</tr>
<tr>
<td>pDT011</td>
<td>bepD promoter region (bp -262 to +13) in pCD366</td>
<td>This study</td>
</tr>
<tr>
<td>pDT012</td>
<td>bepD promoter region (bp -182 to +13) in pCD366</td>
<td>This study</td>
</tr>
<tr>
<td>pDT013</td>
<td>bepD promoter region (bp -102 to +13) in pCD366</td>
<td>This study</td>
</tr>
<tr>
<td>pDT014</td>
<td>virB promoter region (bp -279 to +21) in pCD366</td>
<td>This study</td>
</tr>
<tr>
<td>pDT015</td>
<td>virB promoter region (bp -153 to +21) in pCD366</td>
<td>This study</td>
</tr>
<tr>
<td>pCD353</td>
<td>Km' Tra' mob' oriRIF1010 (IncQ); gfp mut2, fused to pUC19, lacIq</td>
<td>This study</td>
</tr>
</tbody>
</table>
and pDT032,ligated in PCR-blunt II TOPO, digested with BamHI, and inserted into the corresponding site of pCD341.

**Construction of green fluorescent protein (GFP) reporter plasmids.** The 366-bp intergenic upstream region of the virB2 gene (47) and the 333-bp intergenic upstream region of the bepD gene were amplified using primers prT011-prT012 and prT015-prT016, respectively. The terminal EcoRI and BamHI sites were used to insert the fragment in the corresponding sites of pCD346, yielding prVirB-gfp and pPepD-gfp. The truncated promoter vector probes were generated by the same strategy using the primers listed in Table S4 in the supplemental material.

**Construction of expression vectors for recombinant proteins.** The C-terminally His<sub>6</sub>-tagged version of BatR (His<sub>6</sub>-BatR) was generated by amplifying the BatR coding sequence using primers pDT007 and pDT008. Using the flanking NdeI and BamHI sites, the fragment was cloned into the corresponding sites of the expression vector pET15b (+) (Novagen), resulting in pDT009. The gluta-thione S-transferase (GST)-tagged version of the kinase domain of BatG (GST-BatG<sub>1-332</sub>) was constructed by PCR amplification using oligonucleotides pDT026 and pDT027. Using the flanking BamHI and EcoRI sites, the fragment was cloned into the corresponding sites of the expression vector pGEX-2T (GE Healthcare), resulting in pDT020. The C-terminally His<sub>6</sub>-tagged version of VirB8 (His<sub>6</sub>-VirB8) was generated by amplifying VirB8 coding sequence using primers prSH003 and prSH004. Using the flanking NdeI and BamHI sites, the fragment was cloned into the corresponding sites of the expression vector pET15b (+) (Novagen), resulting in pSH007. The C-terminally His<sub>6</sub>-tagged version of BepD (His<sub>6</sub>-BepD) was generated by amplifying BepD coding sequence using primers pPG165 and pPG166. Using the flanking BamHI and NdeI sites, the fragment was cloned into the corresponding sites of the expression vector pET15b (+) (Novagen), resulting in pPG167.

**Infection assay.** EAHy926 cells were grown to confluence in Dulbecco's modified Eagle medium (DMEM) with Glutamax (Gibco Invitrogen) supplemented with 10% fetal calf serum (FCS) (Gibco Invitrogen) in 150-cm<sup>2</sup> cell culture flasks in a humidified atmosphere at 37°C and 5% CO<sub>2</sub>. One hour before infection, the cells were washed with DMEM. The bacteria were grown for 48 h on Columbia agar plates containing 5% defibrinated sheep blood (CBA plates) and harvested in phosphate-buffered saline (PBS) (pH 7.4). M<sup>8</sup> cells were infected at a multiplicity of infection (MOI) of 200. The infected cells were harvested using a cell scraper (Corning), pooled with the supernatant, and centrifuged at 70°C for 3 h into 1 mM IPTG at 37°C and purified as His<sub>6</sub>-BatR. His<sub>6</sub>-BepD was expressed from pDT021 with 4 h of induction using 1 mM IPTG at 37°C and purified as His<sub>6</sub>-BatR. Purified His<sub>6</sub>-VirB8, His<sub>6</sub>-BepD, and His<sub>6</sub>-BatR were loaded on preparative SDS-PAGE. After electrophoresis, the gel was stained with 1 M KCl, the band corresponding to the size of the protein was cut out, and the gel pieces were sent for immunization (Laboratoire d’Hémostologie, Marlite, Belgium).

**Western blot analysis.** To detect His<sub>6</sub>-BepD, sulfatide, pyranoside gels were stained with 12% SDS-PAGE and immunoblotting for the detection of BatR, VirB8, and BepD proteins were performed as follows. B. henselae wild type or the batR mutant was harvested after 48 h of growth on CBA plates or HEC infection (MOI, 200). To recover bacteria from infected HEC, the cells were pretreated with PBS-0.5% saponin (35°C) for 5 min before being harvested. The cells were washed and resuspended in PBS to an optical density at 600 nm (OD<sub>600</sub>) of 16 and resuspended with an equal volume of 0.5% Laemmli buffer and incubated at 4°C in a swinging-bucket rotor. The sample was allowed to auto-fluorimager. For competition with nonlabeled double-stranded oligonucleotides, 1 nmol of purified GST-BatS<sub>1-332</sub> (from pDT020) was induced with 1 mM isopropyl-

**Electrophoretic mobility shift assays (EMSA).** Radiolabeled probes were generated by PCR in the presence of [α-<sup>32</sup>P]dATP or [γ-<sup>32</sup>P]ATP and purified using a nucleotide removal kit (Qiagen). For each probe, a parallel reaction was performed in the absence of radioactive dATP, and the DNA concentration was determined using a NanoDrop ND-100 spectrophotometer (Thermo Scientific). The PCR primers used to generate the different probes are listed in Table S4 in the supplemental material. The PCR was performed with the different probes and the DNA concentration was determined using a NanoDrop ND-100 spectrophotometer (Thermo Scientific). The PCR primers used to generate the different probes are listed in Table S4 in the supplemental material. The reaction mixture was supplemented with 1,000 molar excess of annealed complementary oligonucleotides and incubated at 37°C for 20 min. The gel pieces were sent for immunization (Laboratoire d’Hémostologie, Marlite, Belgium).

**Flow cytometric analysis.** Induction of expression from the virB and bepD promoters was measured as GFP fluorescence by using a FACScan light flow cytometer (Becton Dickinson) with excitation at 488 nm. B. henselae strains carrying a reporter plasmid were streaked from stock (70°C) on CBA plates.
with 30 μg/ml kanamycin and grown in a humidified atmosphere at 35°C and 5% CO₂ for 3 days, followed by restreaking on fresh CBA plates and growth for 48 h. The bacteria were resuspended in PBS, washed, diluted to a final OD₆₀₀ of 0.006 in 1 ml of tester medium, and incubated in 24-well plates in a humidified atmosphere at 37°C and 5% CO₂. To test the pH dependency of GFP induction of B. henselae strains containing pVirB-gfp or pBepD-gfp, M199 was reconstituted from 10× stock solution (Gibco Invitrogen), supplemented with 10% FCS, and buffered with sodium bicarbonate (0.3 g/liter to 8.3 g/liter) to cover a pH range between 6.3 and 8.1 (35°C).

**Phylogenetic analysis of BatR and BatS.** Protein sequences of BatR and BatS and concatenated sequences of BatR/BatS were aligned by using ClustalW implemented in MEGA4 (52). Phylogenetic trees were inferred from the single protein or the concatenated alignments of BatR and BatS by the maximum-likelihood method with PhyML 3.0 (24) using the paralogous phosphate-sensing protein or the concatenated alignments of BatR and BatS by the maximum-likelihood method with PhyML 3.0 (24) using the paralogous phosphate-sensing system PhoB/PhoR as the outgroup. An appropriate substitution model was selected by using the Akaike information criterion of ProtTest (1). We used the model LG + I + G + F for inferring the trees based on BatS, as well as on the concatenated sequences, and the LG + G model for the tree based on BatR.

**Microarray data accession numbers.** The microarray data have been deposited in the microarray database at EBI under accession numbers A-MEXP-644 and A-MEXP-645 for the array design and E-MEXP-2322, E-MEXP-2323, and E-MEXP-2324 for experimental data.

**RESULTS**

**Genome-wide transcriptional profiling during infection of human endothelial cells.** To identify sets of B. henselae genes that are coregulated during human HEC infection, we performed transcriptional profiling using a PCR-based DNA microarray covering 92.3% of the B. henselae protein-encoding genes (1,373/1,488), designed as previously described (4). The results of a time course experiment of Ea.hy926 cell (HEC) infection with B. henselae in DMEM-10% FCS (DMEM) is presented as color-coded concentric circles in Fig. S1 in the supplemental material (circles 1 to 6). We performed a hierarchical gene tree analysis of the 95 genes that fulfilled our criteria for differential expression (see Table S1 and description in the supplemental material). The gene tree presented in Fig. 1 forms four major clusters corresponding to distinct expression profiles: cluster 1 comprises genes displaying progressive but persistent upregulation during the course of infection; cluster 2 comprises genes displaying rapid upregulation after contact with HEC, followed by slower downregulation; cluster 3 comprises genes that become transiently upregulated, followed by downregulation below the initial expression level; and cluster 4 comprises genes that become persistently downregulated during the course of infection (see Table S1 in the supplemental material). Clusters 1 and 2 best matched the expected expression profile for an adaptive response. Strikingly, cluster 1 includes the majority of the 18 genes encoding the VirB T4SS (virB2 to -7, -10, and -11) and its cognate translocated effector proteins (bepC to -E and -G). Figure 2 illustrates the changes in expression at the virB-virD4-bep locus. None of the genes located directly upstream or downstream of the locus displayed differential expression, confirming the specificity of the regulation.

A validation of the microarray data for the genes virB4 and bepD by qRT-PCR is shown in Fig. S2A and B in the supplemental material. This experiment demonstrated congruent data for both methods, except that the fold change values of gene expression derived from the microarray experiment were compressed by a factor of 8 (bepD) to 10 (virB4). This indicates that the amplitude of differential expression might be generally compressed by a factor of 8 (bepD) to 10 (virB4).

![FIG. 1. Identification of B. henselae gene clusters that are distinctly regulated during HEC infection. (A) Hierarchical gene tree cluster analysis of B. henselae genes regulated during HEC infection. Gene tree cluster analysis of a time course experiment of HEC infection (0.25, 6, 24, and 48 h of infection; n = 4) by the wild type (wt) was performed for the 95 genes considered to be regulated (for details, see Materials and Methods). (B) Display of normalized fluorescence intensity ratios of wild-type versus batR mutant bacteria at 48 h of HEC infection in DMEM (n = 6) or 48 h of incubation in M199 (n = 3). The normalized fluorescence ratios calculated for each gene are color coded, with the color scale bar shown at the top. The conditions are indicated at the base of the cluster, and the relationships of the different clades are indicated on the left. Clusters of coregulated genes are separated by horizontal lines. Genes with annotated functions are indicated.](image-url)
underrepresented in the microarray data, as previously reported (40).

**BatR controls the expression of virulence genes during infection.** A previous study on *B. tribocorum* identified the *batR* gene, encoding the putative response regulator of the BatR/BatS TCS, as essential for the colonization of the rat as the natural reservoir host (42). Therefore, we were interested in assessing the involvement of the BatR/BatS TCS in the adaptive response to *B. henselae* during HEC infection. The *batR* gene displayed an expression profile similar to that of cluster 1, although it is not included in the cluster, as it did not pass the quantitative cutoff in fold expression. However, validation of the microarray data for *batR* by qRT-PCR (see Fig. S2C in the supplemental material) showed that the gene is indeed differentially regulated during HEC infection (fold change after 48 h = 11.3 [range, 6.8 to 21.1]). To test whether BatR might be a master regulator of the genes differentially regulated during HEC infection (Fig. 1), we generated an in-frame deletion in the *batR* gene and compared the gene expression profile of this *batR* mutant to the wild-type strain at 48 h of HEC infection in DMEM (see Fig. S1, circle 7, in the supplemental material). In total, we recorded 43 genes that fulfilled our quantitative and statistical criteria for differential regulation (see Table S2 in the supplemental material). Of these genes, 35 were downregulated and 8 were upregulated in the *batR* mutant. Figure 1B and Table S1 in the supplemental material display the expression profiles for the *batR* mutant in comparison to the wild-type for the 95 genes differentially expressed in the wild-type time course experiment. Strikingly, 21 of the 33 genes in cluster 1 (representing the genes gradually upregulated in the microarray time course experiment) are significantly downregulated in the *ΔbatR* strain, with an additional 6 genes that passed the statistical test but not the quantitative cutoff. These results indicate that BatR is indeed a master regulator for this set of genes, including the *virB* operon and the adjacent *bep* genes. The lack of a global effect of *batR* deletion on the regulation of the genes comprising clusters 2 to 4 suggests the involvement of additional regulators and/or the integration of additional stimuli during the infection process.

**The induction of BatR-regulated genes is not dependent on the presence of HEC.** The qRT-PCR data for time points corresponding to the microarray time course experiment displayed in Fig. 3 demonstrate that both *virB4* (representative of the *virB2-virB11* operon) and *bepD* (representative of genes encoding VirB/VirD4-translocated effector proteins) are strongly upregulated during HEC infection in DMEM, while these genes are only marginally upregulated in cell-free DMEM (Fig. 3A and B). These data could suggest that BatR-dependent upregulation of the genes is mediated by the presence of HEC. However, in a similar experiment replacing DMEM with medium 199-10% FCS (M199), we noticed prominent induction of the expression of these two genes in cell-free M199 (38-fold compared to 7-fold induction for *virB4* and 21-fold compared to 3-fold for *bepD* in cell-free DMEM) (Fig. 3C and D). This induction was strictly dependent on BatR, since neither *virB4* nor *bepD* was upregulated in the *ΔbatR* strain, either in the presence of HEC or in cell-free M199-10% FCS (Fig. 3E and F). These results indicate that contact with HEC is not a prerequisite for the induction of these genes. To determine whether the same set of genes would be differentially regulated between the wild-type and the *ΔbatR* strains during HEC infection in cell-free M199 as in DMEM, we determined the gene expression profile of the *batR* mutant compared to that of the wild-type strain after 48 h of M199 induction (see Table S2 in the supplemental material). Indeed, the data presented in Fig. 1B and Fig. S1 in the supplemental material (circle 8) revealed a significant overlap between the 37 genes differentially regulated by BatR in the cell-free medium and the 35 differentially regulated during the infection of endothelial cells, with 22 genes fulfilling both our statistical and quantitative criteria for each condition. Moreover, the combined list of differentially regulated genes for these two conditions demonstrated a high degree of congruency, with 31 of the 43 genes differentially regulated during HEC infection showing a statistical difference in the M199 induction experiment and 29 of the 35 genes differentially regulated in the M199 induction significantly regulated in the HEC infection experiment. In summary, the regulatory cascade triggered by BatR was determined in two different sets of experiments using the wild-type and *ΔbatR* strains grown under both infectious and noninfectious conditions, with a high degree of congruency for both experiments. We thus defined the BatR regulon as the

![FIG. 2. Alteration of *B. henselae* gene expression patterns at the *virB-virD4-bep* locus during HEC infection. (A) Chromosomal organization of the *virB-virD4-bep* locus. (B) Display of the normalized fluorescence intensity ratio of a time-course experiment of HEC infection by the wild type. (C) Display of normalized fluorescence intensity ratios of wild-type versus *batR* mutant bacteria at 48 h of HEC infection in DMEM or 48 h of incubation in M199. The normalized ratios are color coded as in Fig. 1.](image-url)
regulated and 12 genes that are upregulated in the ΔbatR strain.

The batR mutant displays a virulence-attenuated phenotype during HEC infection that can be complemented in trans. To confirm that the global gene expression regulatory phenotype of the batR mutant is due to the deletion of the batR gene itself and is not the result of a coincidental secondary mutation, we complemented the mutant in trans by expressing batR from a low-copy-number plasmid under the control of the tac-lac promoter (ΔbatR-pbatR). qRT-PCR analysis of the BatR-regulated genes virB4 (Fig. 4A) and bepD (Fig. 4B) demonstrated complementation on the mRNA expression level for both genes during HEC infection and cell-independent induction in M199. The marginal downregulation measured for the expression of the virB4 and bepD genes in the ΔbatR strain grown on CBA plates compared to the wild-type (Fig. 4A and B, CBA) strongly supports the hypothesis that these genes are expressed only at a basal level under these conditions. Moreover, the marked induction of the virB4 and bepD genes in the complemented mutant grown on CBA plates in the presence of IPTG demonstrates that overexpression of BatR under these noninducing conditions is enough to bypass the sensory signal necessary for the induction of these genes in the wild type.

Phenotypic restoration of the complemented strain was also demonstrated at the level of host cellular phenotypes provoked by the BatR-regulated genes. Based on the lack of upregulation of the entire virB-virD4-bep gene cluster in the ΔbatR mutant (Fig. 2C), the mutant should be deficient for known T4SS-mediated phenotypes of infected HEC. Indeed, the batR mutant did not provoke the formation of the ring-like F-actin rearrangements that are a characteristic of invasome-mediated uptake of B. henselae (17, 41, 46), while the complemented ΔbatR-pbatR mutant strain triggered these cytoskeletal rearrangements as efficiently as wild-type bacteria (Fig. 4C). Western blot analysis using polyclonal sera raised against VirB5 (a representative for the operon encoding the VirB2-VirB11 proteins) and BepD confirmed the requirement for BatR for the expression of these proteins, since neither VirB5 nor BepD was detectable in the batR mutant, either from bacteria grown on CBA plates or after HEC infection (Fig. 4D).

BatR/BatS constitute a functional TCS. The genetic organization of the batR and batS genes in one locus and the genetic and biochemical data obtained for the closely related TCSs of B. abortus, A. tumefaciens, and S. meliloti suggest that BatS and BatR also constitute a functional TCS. To prove this hypothesis, we designed an in vitro experiment to demonstrate the histidine kinase activity of BatS and the phosphorelay between BatS and BatR. We could demonstrate that the purified cytoplasmic kinase domain of BatS (BatSΔ1-332) autophosphorylates in the presence of [γ-32P]ATP (Fig. 5A, lane 1) and that a phosphotransfer event could take place between the phosphorylated BatSΔ1-332 and BatR (Fig. 5A, lane 2). Conversely, no BatR phosphorylation occurred when BatSΔ1-332 was omitted from the reaction (Fig. 5A, lane 3). A time-course experiment of BatS autophosphorylation showed a gradual increase of radiolabeled BatSΔ1-332 when incubated in the presence of [γ-32P]ATP, reaching 50% phosphorylation after about 15 min (Fig. 5B). A similar time course for the phosphotransfer between prelabeled BatSΔ1-332 and BatR is shown in Fig. 5C. Maximal BatR phosphorylation was reached.

![Graphs](image_url)
after 2 min, followed by a gradual loss of signal. BatR dephosphorylation was not caused by the presence of GST–BatS/H90041-332, since the same loss of signal was observed if the protein was removed by glutathione-Sepharose (data not shown). These results confirm that BatR/BatS constitute a bona fide TCS in B. henselae.

Promoters of the virB operon and the bepD gene are direct targets of BatR. Based on its sequence homology with the OmpR/PhoB subfamily of response regulators, BatR is likely to bind DNA and act as a transcription factor. As BatR regulates the virB operon and the bepD gene, we decided to test whether BatR directly binds to the promoters of these genes. First, promoter-gfp fusions were created for the entire upstream intergenic region of the two genes. A series of truncated derivatives was created to delineate the promoter region, and finally, the direct binding of BatR to these refined promoter regions was assessed by EMSA.

Following initial growth on CBA plates, the wild-type and ∆batR strains carrying pVirB-gfp (bp = 366 to +21) were cultivated for various periods of time in M199, and the GFP-mediated fluorescence intensities of individual bacteria were determined by flow cytometry (Fig. 6A, -366 to +21). The results corroborated those obtained by microarray and qRT-PCR, showing upregulation during M199 induction and a strict BatR-dependent activity for the promoter. Deletion analysis of the virB promoter allowed us to map the BatR regulatory sequence between bp/H11002366 and bp/H11002279 (Fig. 6A). Based on this information, we assessed direct binding of BatR to the virB promoter regions by EMSA and showed a BatR concentration-dependent mobility shift of the radiolabeled PvirB (Fig. 6B, lanes 1 to 4). The EMSA results correlated with the results obtained with the promoter deletion analysis (Fig. 6A), since we observed a similar shift using a shorter fragment covering the putative BatR binding region on PvirB (bp = 366 to -134; Fig. 6B, lanes 9 to 11), but not when using a truncated fragment in which transcription activity was abolished (bp = -153 to +13; Fig. 6A and B), confirming the specificity of the binding assay. To further confirm the specificity of BatR binding to PvirB and to better delineate the BatR binding motif, we designed a set of overlapping oligonucleotides spanning the shortest positive probe in our assay (bp = -366 to -134) and used them as double-stranded unlabeled competitors in our
Results - Research article 1

EMSA experiments in the presence of the radiolabeled probe. Two sets of overlapping competitors prevented the binding of BatR to the radiolabeled probe (Fig. 5C, competitors 3 and 4), whereas the other competitors did not show any interference in the assay (Fig. 6C, competitors 1 and 2 and 5 to 8). These results further confirmed the specificity of the observed shift in the case of the virB promoter and allowed us to locate the BatR binding site between bp 306 and 277 relative to the virB ATG.

The same approach was applied to the bepD promoter (Fig. 7). Analysis of PbepD-gfp (bp −333 to +13) in wild-type and ΔbatR strains grown in M199 showed clear BatR-dependent GFP expression (Fig. 7A, −333 to +13). Notably, only some of the wild-type bacteria exhibited an increase of fluorescence, with the remaining bacteria staying in the uninduced state, indicating a possible bistability phenotype, as described for the Salmonella enterica serovar Typhimurium type III secretion system (2). This biphasic induction pattern was also observed for the virB promoter (Fig. 6A, −366 to +21 and −279 to +21), although most of the bacteria reached the active state under the tested conditions. EMSA demonstrated direct binding of BatR to PbepD (Fig. 7B), in agreement with the promoter deletion analysis (Fig. 7A). Moreover, the competition experiments with unlabeled double-stranded oligonucleotides allowed us to locate the BatR binding site between bp −176 and −137 relative to the bepD ATG.

It should be mentioned that the shifts observed in the autoradiograms presented in Fig. 6B and 7B are not conventional shifts but are more likely to represent DNA-dependent protein oligomerization, since we were not able to resolve this DNA-protein complex into the polyacrylamide gel (the material stayed at the bottom of the well). This phenomenon of oligomerization has already been described for other members of the OmpR subfamily of response regulators (7, 10, 28, 54).

Nevertheless, the clear correlation between the deletion analyses of these two promoters, combined with transcriptional fusion to gfp and the EMSA experiments (including competition experiments with unlabeled oligonucleotides), demonstrates that both the virB operon encoding the VirB T4SS and the bepD gene encoding one of the cognate translocated effectors are direct targets of the BatR/BatS TCS.

The expression of the BatR-regulated virB operon and bepD gene is pH dependent. We next addressed the question of the signal perceived and transduced by the BatR/BatS TCS. In Fig. 3A and B, there is a striking difference in the upregulations of the virB operon and the bepD gene between two different cell culture media (DMEM versus M199) in the absence of HEC. A plausible explanation for this phenomenon could be that it is the result of pH-dependent regulation, since the steady-state pH at 35°C/5% CO2 for DMEM is 7.8 versus 7.35 for M199 in the absence of cells, as a function of the bicarbonate buffer system (3.7 g/liter versus 2.2 g/liter bicarbonate, respectively). This inherent property of the media was balanced in the presence of HEC by the progressive acidification observed during the culture of the cells, reaching steady-state pHs of 7.2 in M199 and 7.4 in DMEM.

We hypothesized that BatS may represent a pH sensor. To test this hypothesis, we used the wild-type and ΔbatR strains carrying pVirB-gfp or pBepD-gfp as biosensor strains and cultivated them for various periods in M199-10% FCS adjusted to different pHs with appropriate concentrations of bicarbonate. The fluorescence histograms for a representative time course experiment performed at acidic (6.3), neutral (7.35), or basic (8.0) pH are illustrated in Fig. 8A and B. The pH-dependent induction of the two promoters investigated is summarized in Fig. 9. In the wild-type strain, the two promoters displayed similar pH-dependent upregulation over a narrow range of neutral to slightly basic pHs (pH 7.0 to 7.8 for PvirB and pH 7.3 to 7.9 for PpbeD), whereas no induction was detectable for the batR mutant under any of the tested conditions. In the wild type, the virB promoter responded earlier (as early as 24 h) than the bepD promoter (at 48 h), in agreement with the qRT-PCR results presented in Fig. 3. The minor differences in the measured pH optima of the tested promoters are in the range of the experimental variation (error bars in Fig. 9A and B) and therefore are not statistically significant. While none of the promoters was induced at acidic pH (Fig. 8A and B), the induction of both PvirB and PpbeD at basic pH was not fully abolished but was greatly reduced. In a control experiment, both wild-type B. henselae and ΔbatR strains carrying an IPTG-inducible tac-lac promoter fusion to gfp showed upregulation to the same extent at all three tested pH values, indicating that none of these conditions interfered with transcription or translation in B. henselae (Fig. 8C and D). Plating of the bacteria incubated at different pHs revealed only minor effects on growth for both wild-type and batR strains when acidic (pH 6.4) and neutral (pH 7.6) conditions were compared.
For the basic (pH 8.1) condition, however, a drop in CFU for the wild-type but not the batR strain was observed after 48 h of incubation. These results confirmed that the differences in induction of both $P_virB$ and $P_bepD$ between acidic and neutral pH were not a consequence of different survival of the wild-type bacteria. However, this cannot be excluded for induction at basic pH. 

Figure 9C and D demonstrates the same pH-dependent regulation of $virB4$ and $bepD$ using qRT-PCR, validating the results obtained with the biosensor strains. In summary, we demonstrated that the BatR-mediated upregulation of $P_virB$ and $P_bepD$ is pH dependent, with optimal activation around physiological pH. Therefore, we postulate that BatS constitutes a pH sensor that, together with its cognate response regulator, 

(see Fig. S3 in the supplemental material). For the basic (pH 8.1) condition, however, a drop in CFU for the wild-type but not the batR strain was observed after 48 h of incubation. These results confirmed that the differences in induction of both $P_virB$ and $P_bepD$ between acidic and neutral pH were not a consequence of different survival of the wild-type bacteria. However, this cannot be excluded for induction at basic pH.
BatR, mediates the adaptive response of *B. henselae* during hemotropic infection of HEC.

The **BatR regulon: conserved regulators and mobile target genes.** Finally, we were interested in examining the extent to which the BatR/BatS system has coevolved with its regulon. To this end, we constructed a phylogenic tree of the concatenated sequences of BatR/BatS orthologous proteins in the alphaproteobacteria using the maximum-likelihood method and estimated the congruence of the resulting tree with previously inferred species tree topology (43, 58). The results showed that

---

**FIG. 7.** BatR binding to the *bepD* promoter. (A) Deletion analysis of the 333-bp intergenic region upstream from the *bepD* gene by promoter-GFP reporter probe vector in wild-type (open) and Δ*batR* (shaded) strains. Bacteria were incubated in M199-10% FCS (pH 7.4) for 48 h and analyzed by flow cytometry. The positions of the probes relative to ATG of the *bepD* gene are indicated. (B) Specific binding of the purified BatR protein to the *bepD* promoter demonstrated by EMSA. Three radiolabeled DNA probes (2 to 4 fmol) spanning the full-length *PbepD* or subfragments were incubated with 0, 0.8, 1.6, or 3.2 μg of purified BatR protein. The positions of the probes relative to the ATG of the *bepD* gene are indicated. (C) EMSA of the 102-bp fragment of the *bepD* promoter in the presence of unlabeled competitors. Two femtomoles of radiolabeled probe was incubated with either no (−) or 2 μg (+) purified BatR protein in the absence (−) or presence of 1,000× molar excess of double-stranded competitor (1 to 6). The position of the competitor relative to the probe is displayed.
Results - Research article I

A

B

C

D
the BatR/BatS tree was consistent with the topology of the species tree for all nodes with significant bootstrap support (Fig. 10), indicating strict vertical inheritance of the BatR/BatS system. Similar tree topologies were obtained when only the BatR or BatS protein sequence was used (see Fig. S4 in the supplemental material), confirming coinheritance of the two components. To investigate the conservation among the alpha-proteobacteria of the 58 genes representing the BatR regulon (as determined by our transcriptional-profiling analysis), we analyzed their presence in the species included in the BatR/BatS gene tree. To this end, we used BlastP and determined the degree of sequence identity for each pair of homologs by calculating their conservation scores (CS) (see Table S3 in the supplemental material) (38). This analysis revealed that only 15 of these 58 genes have homologs within most alpha-proteobacteria (Fig. 10, Nod1; see Table S2 in the supplemental material), only 6 of which are relatively well conserved in the analyzed species (see Table S3 in the supplemental material). Overall, these well-conserved genes essentially encode proteins with housekeeping functions (e.g., the chaperonin genes groEL-groES, the serine protease gene htrA1, and the GMP synthase gene guaA) or with a role in transcription regulation (the sensor histidine kinase genes BH04790 and envZ; the response regulator genes BH04780, BH13850, and batR; and the transcription regulator gene rosR). Conservation analysis of the remaining genes showed that, except for six genes conserved in several, but not all, of the Rhizobiales (Fig. 10, node 2), including hbpA and hbpB, encoding hemin-binding proteins, the remaining 37 genes appeared to be unique to the bartonellae or their presence in the bartonellae resulted from horizontal gene transfer, as in the case of the virB operon (Fig. 10, nodes 3 and 4) (42). Moreover, conservation analysis of the

FIG. 8. pH-dependent activation of PvirB and PbepD. (A and B) Fluorescence histograms of the wild-type (open) and ΔbatR (shaded) strains harboring a virB (A) or bepD (B) promoter fusion to gfp for a representative time course experiment performed at acidic (pH 6.3), neutral (pH 7.35), or basic (pH 8.0) pH. The thresholds used for the scoring of GFP-positive bacteria in Fig. 9 are indicated. (C and D) The expression of an IPTG-inducible tac-lac promoter fusion to gfp is not affected by a change in pH. Shown are fluorescence histograms of the wild-type (C) and ΔbatR (D) strains harboring a tac-lac promoter fusion to gfp for a representative time course experiment performed at acidic (pH 6.3), neutral (pH 7.35), or basic (pH 8.0) pH in the presence (open) or absence (shaded) of 0.5 mM IPTG.
37 Bartonella-specific genes revealed that 27 were absent from the genome of B. bacilliformis and were therefore restricted to the radiating lineage of Bartonella (Fig. 10, node 4) (42). In summary, conservation analysis of the BatR regulon within the alphaproteobacteria suggested that a highly conserved monitoring system modulates the expression of a rapidly evolving set of target genes to adjust the structure and physiology of the bacterium to its environment.

DISCUSSION

In this study, we have addressed the coordinated response orchestrated by B. henselae to successfully invade and colonize HEC using transcriptional profiling. Specific attention has been paid to the BatR/BatS TCS and its role in the regulation of the VirB T4SS and its translocated effector proteins. The interaction of B. henselae with HEC has been extensively studied (14, 15, 17, 19, 46), and substantial progress has been achieved toward the understanding of the role of the VirB T4SS and its secreted Beps in the mediation of this interaction (41, 44, 45, 49). However, the regulation mechanism(s) controlling the expression of these virulence factors and the global response of the bacteria during HEC interaction remain elusive.

Adaptive response during HEC infection. In this paper, we showed that the virB operon and the bep genes are induced during endothelial cell infection, together with a set of coregulated genes (Fig. 1, cluster 1). The progressive but persistent upregulation of these genes best fit the expected gene expression profile enabling an adaptive response. The results obtained from the analysis of a batR mutant revealed that most of the genes found in this cluster are under the positive control of the response regulator BatR. Among these genes, we also found two of the four B. henselae heme receptor gene family members, hbpB and hbpC, whereas hbpA and hbpD were found to be downregulated during infection (Fig. 1, cluster 4). Different expression patterns for the members of this gene family have been described in B. quintana (5), where hbpA and hbpD were expressed at low heme concentrations while hbpB and hbpC were expressed at high heme concentrations. We described a second cluster of genes displaying rapid upregulation after contact with HEC, followed by slower downregulation during the time course of infection (Fig. 1, cluster 2). Different expression patterns for the members of this gene family have been reported in B. quintana (5), where hbpA and hbpD were expressed at low heme concentrations while hbpB and hbpC were expressed at high heme concentrations. We described a second cluster of genes displaying rapid upregulation after contact with HEC, followed by slower downregulation during the time course of infection (Fig. 1, cluster 2).
fied as the major regulatory factor for transcriptional osmotic regulation in E. coli (34, 53). Furthermore, ompR was shown to be involved in the virulence and survival of Brucella melitensis during macrophage infection (59), and a recent study suggested the involvement of ompR in the ability of B. henselae to invade HEC (23). Gene cluster 2 thus seems to reflect the B. henselae response to the drastic change in environment encountered at the onset of the in vitro infection and may comprise factors important for the establishment of contact with the HEC. Since the deletion of batR did not result in any significant alteration in gene expression for this gene cluster, other transcriptional regulators must be involved in their regulation, and the OmpR/EnvZ TCS represents a likely candidate to fulfill this function. The third cluster of genes upregulated during HEC infection and displaying transient upregulation mostly contains genes within a Bartonella-specific island including phage genes (21). The relevance of this transient upregulation is currently unknown; however, the precise timing of this regulation suggests a role in the early stage of interaction with the endothelial cells or a response to the process. Overall, we have presented here the first comprehensive analysis of B. henselae transcriptional response to HEC infection, which can be grouped into four main clusters of coregulated genes. The three clusters showing a complementary upregulation pattern may represent the sequential physiological changes and/or critical factors required to successfully invade and colonize the HEC.

BatR-regulated genes. Bacterial TCSs have been associated with different strategies to achieve niche adaptation. A widespread strategy relies on the acquisition of new TCSs by horizontal gene transfer and/or the lineage-specific expansion of existing gene families, as revealed by the analysis of nearly 5,000 histidine kinase genes in 200 bacterial genomes (3). In comparison to these broader studies, we have focused here on one particular TCS, for which we show that both the histidine kinase and the response regulator have evolved by vertical descent over a long evolutionary distance. The conserved mode of evolution for the sensing components of the system is contrasted by rapid diversification of the regulon involved in formulating the response, as revealed by a striking number of genus-specific genes. The rapid diversification of genes regulated by conserved regulatory proteins has been reported as an alternative strategy to achieve niche adaptation (37), with the regulon of orthologous transcription factors being typically divided into shared target genes and species-specific genes. The shared genes of a regulon allow the bacteria to cope with the environmental change that activates the regulon and to control the amount of the active form of the transcription factor, whereas the species-specific targets help each species to proliferate in the particular niches in which they live (37), as exemplified by a recent study on the evolution of the PhoP/PhoQ regulon in the enterobacteria (38). Based on our conservation analysis, the BatR regulon seems to respect this categorization, as it consists of a small set of genes found in most of the alphaproteobacteria that encode either housekeeping (e.g., groEL, groES, guaA, and rimN) or transcriptional regulators, and a large subset of genus-specific genes, most of which are absent in B. bacilliformis and thus are restricted to the radiating lineage of the bartonellae. Prime among these genes we found the virB locus encoding the VirB T4SS and genes of the flanking bep cluster, encoding the T4SS-translocated effector proteins. We also found two components of the Trw T4SS (trwN and trwJ1), the second T4SS of B. henselae, and one of its regulatory proteins (korA). The Trw T4SS was shown to be essential for intraerythrocytic parasitism of B. tribocorum in the rat infection model (42, 50) but appears to be dispensable for the infection of HEC (15). We also found hecB, a horizontally acquired gene present in multiple copies in the B. henselae genome, that together with fhaC forms a two-partner secretion system. The fhaC-hecB gene products mediate the transport of filamentous hemagglutinin encoded by fhaB (4). All of these genes appear to be under the positive control of BatR, and strikingly, most of them were more than 2-fold upregulated after 48 h of HEC infection (see Table S2 in the supplemental material), supporting the idea that these genes may represent important factors for B. henselae to colonize HEC. Recently, the ChvI regulon of S. mellotii was described based on the differential gene expression between a partial-loss-of-function chvI mutant and a gain-of-function chvI mutant compared to wild-type expression (9). Strikingly, only 4 of these 59 genes appear to be conserved in B. henselae (data not shown), and none of them is under the control of BatR. Furthermore, none of the three direct transcriptional target genes described for S. mellotii ChvI (9) share homologs with B. henselae, emphasizing the rapid diversification of the genes under the control of this conserved TCS, as demonstrated by our conservation analysis of the BatR regulon. Taken together, the BatR regulon consists of a small core of genes that are conserved among the alphaproteobacteria (although these genes do not appear to be under the control of the S. mellotii homolog) and a large proportion of genes specific to the radiating lineage of the bartonellae. This second group encodes a panel of factors critical for the host cell interaction, as exemplified by the VirB T4SS and its secreted substrates.

The activation of BatR/BatS-regulated genes is pH dependent. Our knowledge of the importance of TCSs in enabling mutualistic or pathogenic host interactions by diverse alphaproteobacteria (29, 32, 60) is contrasted by a sparse knowledge of the environmental signals these systems sense, reflecting the diverse host-associated microenvironments the different alphaproteobacteria colonize. Reporter fusions of the virB and bepD promoters to gfp revealed a pH-dependent activation of these BatR-regulated genes, with an optimum at the physiological pH of mammalian blood (pH 7.4). However, contact with HEC did not appear to be necessary for this response, suggesting that pH sensing is one of the mechanisms used by B. henselae to discriminate the host environment from the arthropod vector. This pH dependency resembles the pH-sensitive system ChvG/ChvI in A. tumefaciens. This plant-associated species infects wounded plant cells and genetically transforms them by the transfer of the so-called T-DNA via the T4SS VirB/VirD4. Plants release acids at sites of wounding, and the external acidification is sensed by the BatS orthologue ChvG. This condition can be mimicked under laboratory conditions by acidifying the growth medium to pH 5.5 (29). Under these conditions, chvI was shown to be upregulated, suggesting positive-feedback regulation (61), which is also observed for batR during HEC infection. Although the transcriptional analysis of the A. tumefaciens response to acid conditions has been analyzed in detail (61), the ChvI regulon has not been resolved to

Results - Research article I

ADAPTIVE TRANSCRIPTIONAL RESPONSE OF B. HENSELAE

VOL. 192, 2010

53

55
date. It is notable that T4SSs were also among our set of genes regulated by BatR/BatS. However, previous studies have shown the T4SSs to be frequently horizontally transmitted (22) and that the two operons shown the T4SS to be frequently horizontally transmitted (22).

It is notable that T4SSs were also among our set of genes regulated by BatR/BatS. However, previous studies have shown the T4SSs to be frequently horizontally transmitted (22) and that the two operons shown the T4SS to be frequently horizontally transmitted (22).

ACKNOWLEDGMENTS

We thank Arturo Mulliainen and Jacob Malone for critical reading of the manuscript. We thank Marc Folcher for helpful suggestions regarding the EMSA experiments.

This work was supported by grant 3100AO-109925 from the Swiss National Science Foundation, grant 55005501 from the International Research Scholar program of the Howard Hughes Medical Institute, and grant 15RT-0 126008 (InfectX) in the framework of the SystemsXChiw Initiative for Systems Biology (to C.D.), as well as by grants from the Swedish Research Council, the Göran Gustafsson Foundation, and the Swedish Foundation for Strategic Research (to S.G.E.A.).

REFERENCES


3.2. RESEARCH ARTICLE II (published)

Directed shotgun proteomics guided by saturated RNA-seq identifies a complete expressed prokaryotic proteome

Ulrich Omasits, Maxime Quebatte, Daniel J. Stekhoven, Claudia Fortes, Bernd Roschitzki, Mark D. Robinson, Christoph Dehio, and Christian H. Ahrens


3.2.1. Statement of the own participation

My contribution to the *research article II* primarily consisted in the establishment of *in vitro* induction conditions that allowed the expression of the BatR regulon, including the VirB/D4 T4SS and its secreted effectors in the absence of eukaryotic host cells and the production of milligram range of protein extracts. I created and validated the strains used in this study, prepared and tested all extracts analyzed in this study, including the RNA samples used for RNA-seq analysis. I took part to the development of the experimental design and contributed to data analysis and to the writing of the manuscript with a special focus on the sections related to the biological implications of the study.
Directed shotgun proteomics guided by saturated RNA-seq identifies a complete expressed prokaryotic proteome

Ulrich Omasits,1,2 Maxime Quebatte,3 Daniel J. Stekhoven,1 Claudia Fortes,4 Bernd Roschitzki,4 Mark D. Robinson,1,5 Christoph Dehio,3 and Christian H. Ahrens1,6,7

1 Quantitative Model Organism Proteomics, Institute of Molecular Life Sciences, University of Zurich, 8057 Zurich, Switzerland; 2 Zurich Life Sciences Graduate School Program in Systems Biology, 8057 Zurich, Switzerland; 3 Biozentrum Basel, University of Basel, 4056 Basel, Switzerland; 4 Functional Genomics Center Zurich, ETH & University of Zurich, 8057 Zurich, Switzerland; 5 SIB Swiss Institute of Bioinformatics, University of Zurich, 8057 Zurich, Switzerland

Prokaryotes, due to their moderate complexity, are particularly amenable to the comprehensive identification of the protein repertoire expressed under different conditions. We applied a generic strategy to identify a complete expressed prokaryotic proteome, which is based on the analysis of RNA and proteins extracted from matched samples. Saturated transcriptome profiling by RNA-seq provided an endpoint estimate of the protein-coding genes expressed under two conditions which mimic the interaction of Bartonella henselae with its mammalian host. Directed shotgun proteomics experiments were carried out on four subcellular fractions. By specifically targeting proteins which are short, basic, low abundant, and membrane localized, we could eliminate their initial underrepresentation compared to the estimated endpoint. A total of 1250 proteins were identified with an estimated false discovery rate below 1%. This represents 85% of all distinct annotated proteins and ~90% of the expressed protein-coding genes. Genes that were detected at the transcript but not protein level, were found to be highly enriched in several genomic islands. Furthermore, genes that lacked an ortholog and a functional annotation were not detected at the protein level; these may represent examples of overprediction in genome annotations. A dramatic membrane proteome reorganization was observed, including differential regulation of autotransporters, adhesins, and hemin binding proteins. Particularly noteworthy was the complete membrane proteome coverage, which included expression of all members of the VirB/D4 type IV secretion system, a key virulence factor.

A major goal of the post-genome era is to understand how expression of the functional elements encoded by a genome is orchestrated to allow an organism to develop and adapt to life under varying conditions. Transcriptomics and proteomics technologies both provide important and complementary insights: The former allow researchers to generate global quantitative gene expression profiles and to study gene regulatory aspects like the impact of short RNAs. However, due to the varying correlation of transcriptomics and proteomics data reported in the literature (de Godoy et al. 2008; de Sousa Abreu et al. 2009; Maier et al. 2011; Marguerat et al. 2012), the direct measurement of protein expression levels is often desirable. For certain aspects, proteomics data can provide more informative and accurate data, as it reflects the effects of other important regulatory processes like protein translation rates and protein stability (Schwanhausser et al. 2011). Furthermore, proteomics provides unique functional insights including post-translational modifications, subcellular localization information, and identification of interaction partners of proteins.

Due to enormous advances in mass spectrometry instrumentation, biochemical fractionation methods, and computational approaches, proteomics has matured into a state where the description of complete proteomes expressed in a specific condition is within reach. So far, only one study has claimed the identification of a complete proteome expressed in haploid and diploid baker's yeast (de Godoy et al. 2008), while extensive proteome coverage has been reported for several prokaryotes (Jaffe et al. 2004; Becher et al. 2009; Malmstrom et al. 2009) and archaea (Giannone et al. 2011). Describing extensive proteome maps under different conditions with a discovery proteomics approach is an important first step in defining the protein expression landscape for an organism and facilitates a subsequent shift away from the discovery mode to a measurement or scoring mode (Kuster et al. 2005; Ahrens et al. 2010).

Due to the lower transcriptome and proteome complexity compared to eukaryotes, an exhaustive discovery proteomics approach is particularly amenable for prokaryotes. We describe here a generic strategy to achieve an essentially complete coverage of a prokaryotic proteome expressed under specific conditions. Key elements of the strategy are the parallel extraction of RNA and...
Results - Research article II

Covering complete expressed prokaryotic proteomes

protein from matched samples, and a saturated transcriptome analysis by RNA-seq (Wang et al. 2009). This in turn allows the generation of a condition-specific endpoint estimate of the number of actively transcribed protein-coding genes, which is a more appropriate estimate than considering all annotated protein-coding genes. A combination of experimental and computational strategies is then used to dig very deep into the proteome.

We apply the strategy to two conditions that mimic the changing environment encountered by Bartonella henselae upon transfer by its arthropod vector into its mammalian host. The Gram-negative α-proteobacterium B. henselae is a hemotropo, zoonotic pathogen that frequently causes cut scratch disease in immuno-competent humans, as well as bacteraemia, endocarditis, and vasoproliferative lesions in immuno-compromised patients. Members of the genus Bartonella are considered re-emerging pathogens and are primarily being studied as models for host-pathogen interaction (Harms and Dehio 2012). A particular emphasis was put on achieving an extensive coverage of the important membrane proteome (Savas et al. 2011). Membrane proteins carry out essential functions as transporters, enzymes, receptors to sense and transmit signals, and adhesion molecules. In light of the resurgence of infectious diseases, membrane proteins are, furthermore, prime candidates for the development of urgently needed novel anti-infectives (Norby et al. 2005).

Relying on a very stringent false discovery rate (FDR) cutoff, we were able to identify 1250 of the 1467 annotated distinct B. henselae proteins, i.e., a proteome coverage of 85%. Several lines of evidence indicated that we have exhaustively measured the expressed proteome and can claim to have identified a complete membrane proteome. This included expression evidence—to our knowledge for the first time—for all protein components of a bacterial type IV secretion system (T4SS) which spans the inner and outer bacterial cell membranes.

Results and Discussion

Model system to explore complete proteome coverage

We chose B. henselae as a model system for several reasons: (1) Its relatively small genome (1.93 Mbp) comprises 1488 predicted protein-coding genes (Alsmark et al. 2004); (2) it is a facultative intracellular pathogen that can be grown in pure culture; (3) protocols for subcellular fractionation have been described (Rhomberg et al. 2004); and (4) in vitro conditions that mimic the pH-dependent induction of virulence genes required for the successful interaction with host endothelial cells, the likely primary niche for B. henselae (Harms and Dehio 2012), have been established (Quebatte et al. 2010). The availability of a model system that eliminates the need for coculture with human endothelial cells is critical to achieve complete coverage of an expressed proteome.

Our in vitro model system relies on the induction of the transcription factor BatR (BH00620) that is essential for the pathogenicity of B. henselae (Quebatte et al. 2010) (for details, see Supplemental Methods; Supplemental Tables S1, S2). In the absence of IPTG (uninduced condition), the batR regulon is not induced, resembling the situation encountered in the arthropod midgut. In contrast, batR expression is up-regulated in the induced condition, resulting in a marked induction of the batR regulon, including the VirB/D4 type IV secretion system (T4SS), which is required for infection of endothelial cells (Schulein and Dehio 2002). This state mimics the environment encountered by bacteria in the mammalian host.

A generic strategy for complete proteome coverage by discovery proteomics

We rely on our previous definition of complete proteome coverage, i.e., having identified protein expression evidence for the annotated

---

**Figure 1.** Overview of the complete expressed proteome discovery workflow. (A) Extraction of RNA and proteins from matched samples, transcriptome analysis. Total RNA and proteins were extracted in parallel from bacteria grown either under uninduced or induced conditions (schematically shown by black knobs representing the VirB/D4 T4SS). Protein extracts were subfractionated into cytoplasmic (Cyt), total membrane (TM), inner (IM), and outer membrane (OM) fractions. To estimate an upper bound for the number of actively transcribed protein-coding genes, the transcriptome was sequenced to saturation using RNA-seq. (B) Analysis-driven experimentation (ADE). In a first pilot phase, samples are analyzed by LC-MS/MS. Underrepresented proteome areas are identified based on a statistical analysis comparing experimentally identified proteins to all expressed proteins (the estimated RNA-seq endpoint indicated by the orange dashed line within an error envelope). All distinct annotated proteins are indicated by the black dashed line. Subsequently, these areas are investigated by targeted experiments, aiming to overcome the saturation trend. (C) Integrative data analysis. Data from the expressed proteome are integrated with genomic, transcriptomics, orthology, and other information to enable further analyses.
protein-coding genes actively transcribed in a given state (Ahrens et al. 2010). A recent proteogenomics study of 46 prokaryotes indicated that, on average, only 0.8% protein-coding genes were missed in the original genome annotations (Venter et al. 2011), justifying our focus on the reference genome. Our strategy to achieve as complete as possible coverage of the expressed proteome of a prokaryote consists of three stages.

In a first stage, RNA and proteins are extracted from identical samples, and whole transcriptome libraries are sequenced to saturation by RNA-seq (Fig. 1A). Thereby, the number of protein-coding genes actively transcribed in a given state can be estimated, shown here for the sum of protein-coding genes expressed in the uninduced and induced condition (orange dashed line, Fig. 1B). Based on such an optimal endpoint estimate, in a second stage, several pilot experiments are performed on cytoplasmic and total membrane fractions of the respective conditions. Following a statistical comparison of the pilot phase proteome (green line, Fig. 1B) to the predicted endpoint, areas of underrepresentation can be targeted by the analysis-driven experimentation (ADE) feedback-loop strategy (Brunner et al. 2007), which can help to overcome the premature saturation of distinct protein identifications and sequence deeper into the expressed proteome (blue lines, Fig. 1B). In a third stage, evidence is presented that virtually no biases remain when comparing protein parameters of all identified proteins to those called actively expressed, justifying the claim to have identified a complete proteome expressed in a specific condition. Analysis of such a data set is expected to provide novel insights regarding the achievable membrane proteome coverage, differential protein expression, and evolutionary conservation and genome structure (Fig. 1C).

Transcriptome exploration by RNA-seq

We relied on RNA-seq (Wang et al. 2009) primarily to generate an endpoint estimate for the number of expressed protein-coding genes. Whole transcriptome libraries of two biological replicates per condition were generated using a protocol that enriches for mRNA transcripts (see Methods). We sequenced very deep into the transcriptome and obtained 55–87 million single end 50-mer reads per sample. Of these, 10.7–26.7 million reads mapped unambiguously, while the vast majority of remaining reads originated from multiple-copy rRNA genes (see Methods; Supplemental Table S3). Reads per kilobase per million (RPKM) values (Mortazavi et al. 2008) showed very high concordance of the biological replicates ($r > 0.97$) (Supplemental Fig. S1).

To estimate how many protein-coding genes are actively expressed in the two conditions, we plotted the number of distinct expressed protein-coding ORFs as a function of the sum of uniquely mapping reads. We required at least five distinct reads within a 50-nt window of the 5' end to deem a protein-coding gene actively expressed (Supplemental Fig. S2), a cutoff similar to that used by Wang et al. (2009). Saturation is characterized graphically through flattening of the curves as the number of reads increases. Due to the asymptotic nature of saturation curves, reaching complete coverage is theoretically only possible with infinite effort. Therefore, we define saturation as the number of discoveries from where, based on nonlinear modeling and extrapolation, a doubling of effort is expected to increase the number of discoveries only marginally. Figure 2A indicates that doubling the number of reads would increase the number of detected protein-coding genes by <3.5% for sample uninduced2 and by ~1% for induced2. Therefore, our analysis indicated that the transcriptome was sequenced to saturation (Fig. 2A). We acknowledge that different library preparations might potentially identify additional genes and that very low abundance transcripts (and proteins) expressed in only a few cells of the population may not be identified with this approach.

We also plotted the density of the RPKM values in order to assess the distribution of transcription levels for all annotated protein-coding genes: The resulting bimodal graph suggested that, under the conditions studied, not all protein-coding genes are actively expressed; RPKM = 10 might be considered a conservative lower cutoff (Fig. 2B). The average RPKM values for members of the virB/D4 operon in condition uninduced2 (30), where the operon is expected to be expressed at low levels, versus induced2 (160) support this observation.

Based on the combined thresholds, 1353 protein-coding genes were expressed in the two conditions (uninduced 1254 and induced 1349). An inter-replicate analysis revealed >95% overlap of the expressed protein-coding genes (Supplemental Table S4). We include an error envelope of ±2.5% to account for uncertainty in the thresholds (Fig. 3A).

Extended proteome coverage strategy: Experimental and computational approaches

Our experimental strategy to reach very deep into the proteome relied on four elements: first, we used a combination of subcellular fractionation and additional biochemical fractionation regimens to
reduce the overall sample complexity. A measure that had been key to describing the complete expressed proteome of baker’s yeast (de Godoy et al. 2008). Second, an exclusion list approach (Kristensen et al. 2004) was applied, which helped to identify a significant amount of low-abundance proteins (Supplemental Fig. S5). Third, we relied on the analysis-driven experimentation feedback-loop strategy (Fig. 1B; Brunner et al. 2007) to target underrepresented areas of the proteome and overcome premature saturation. Finally, for all membrane-derived fractions, we used chymotrypsin in addition to trypsin, thereby maximizing the per-protein sequence coverage and the overall membrane proteome coverage (Fischer et al. 2006).

In terms of computational approaches, we combined results from two database search engines, Mascot Percolator (Brosch et al. 2009) and MS-GF+, an updated version of MS-GFDB (Kim et al. 2010; see Methods), which employs the generating function approach (Kim et al. 2008) to compute statistical significance of peptide identifications (spectra l probabilities). Based on these spectral probabilities or the target-decoy option, one can estimate and stringently control the FDR rate, a critical step for a complete proteome discovery project. Otherwise, lower quality peptide spectrum matches (PSMs) will start to accumulate false-positive peptide evidence for proteins in a random fashion (Reiter et al. 2009). In addition, the error propagates and increases from spectra to peptides and proteins (Nesvizhskii 2010); a PSM level FDR of 1% can correspond to a protein level FDR of 8%–11% (Balgley et al. 2007). We, therefore, chose a very stringent PSM FDR cutoff of 0.01%, allowing us to report protein identifications with an FDR below 1% (see below).

Identification of the complete expressed \textit{B. henselae} proteome

The induction of \textit{batR} and \textit{virB/D4} T4SS expression was more pronounced for the sample pair uninduced2/induced2 than for its biological replicate based on the RNA-seq data. Subcellular fractions from this sample pair (i.e., cytoplasmic [Cyt], total membrane [TM], inner [IM] and outer membrane [OM] fractions) were thus analyzed in detail using different biochemical fractionations (see Methods; Fig. 1A).

We first measured the Cyt and TM fractions of both conditions using OFFGEL electrophoresis at the protein level (OGEprot). When requiring at least two independent PSMs to identify a protein, 924 distinct proteins were identified in four experiments, i.e., 63% of all 1467 distinct annotated proteins or 68% compared to the RNA-seq endpoint estimate of 1353 expressed proteins (Fig. 3A). Analysis of the IM fractions from uninduced and induced condition (IMu/i) and the OM fractions contributed 130,000 additional PSMs (72% more PSMs) but only added 22 previously not identified proteins (Fig. 3A), indicating that we were already in the saturation phase. We fitted a saturation curve to the eight OGEprot experiments, which shows the anticipated trend of further protein identifications assuming no change in the experimental approach, and also calculated confidence intervals (see Methods; Fig. 3A). Carrying out further OGEprot experiments is predicted to lead only to a handful of new protein identifications.

Instead, we relied on the ADE strategy to break the saturation trend. We computed several physicochemical parameters for all \textit{B. henselae} proteins (see Supplemental Methods). The statistical comparison of the parameters of 946 proteins identified in pilot experiments using OGEprot is contrasted to that of all expressed proteins (orange) and those identified in pilot experiments using our ADE approach (green). The most important aspects of over- or underrepresentation can be seen on the abscissa; they indicate that the targeted experiments successfully add new protein identifications in areas of the proteome that were underrepresented in the pilot experiments. For details on the density estimation and the bootstrap confidence bands (shaded areas), see Supplemental Methods.

Figure 3. Overcoming the saturation of protein identifications using ADE-guided shotgun proteomics. (A) Increase of distinct identified proteins given the number of PSMs observed in different experiments. We fitted an exponential curve (see Methods) to all experiments for a given biochemical fractionation in order to find a saturation limit (see colored numbers on the right-hand side). We also approximated confidence bands for the fitted points (thin lines; see Methods). The black dashed line at the top signifies the total number of distinct \textit{B. henselae} proteins (1467); the orange dashed line below represents the estimated RNA-seq endpoint of expressed distinct proteins (1353) including a ±2.5%-error envelope (orange shaded area). (B) Density estimates of four physicochemical protein parameters for different protein subsets. The parameter density for proteins newly identified by the ADE approach is contrasted to that of all expressed proteins (orange) and those identified in pilot experiments using OGEprot (green). The most important aspects of over- or underrepresentation can be seen on the abscissa; they indicate that the targeted experiments successfully add new protein identifications in areas of the proteome that were underrepresented in the pilot experiments. For details on the density estimation and the bootstrap confidence bands (shaded areas), see Supplemental Methods.
by OGEPep in the pilot phase versus the RNA-seq endpoint estimate of 1353 expressed proteins in both conditions provided evidence for a significant underrepresentation of short, low-abundance, basic, and hydrophobic proteins. These areas of the proteome were subsequently targeted by specific experimental approaches (see Supplemental Methods). Underrepresentation with respect to length was targeted using size exclusion chromatography (gel filtration) (Brunner et al. 2007). These experiments added 83 new protein identifications compared to the OGEPep pilot phase (Fig. 3A, blue color). The enrichment for shorter proteins can be appreciated in the upper left panel of Figure 3B. Low-abundance proteins were targeted using ProteoMiner (Guerrier et al. 2008; Fonslow et al. 2011). These experiments (Fig. 3A, gray) helped to identify 42 additional proteins, which were preferentially lower-abundance proteins as evidenced from the density distribution of their Codon Adaptation Index (CAI) values (Fig. 3B, upper right panel; Sharp and Li 1987). Basic and membrane-localized proteins were targeted using OFFGEL electrophoresis at the peptide level (OGEPep). The 285 proteins newly added by the OGEPep experiments (Fig. 3A, red) were highly enriched for basic proteins (Fig. 3B, lower left panel) and membrane proteins (with a high grand-average hydropathicity [gravy] value) (Fig. 3B, lower right panel).

Overall, we identified 1250 distinct proteins requiring at least two PSMs per protein (Supplemental Fig. S3) and only considering peptides that unambiguously identify one bacterial protein (Table 1; Qeli and Ahrens 2010), i.e., 85% of the 1467 distinct protein sequences. The FDRs at the PSM, peptide, and protein level are below 0.01%, 0.1%, and 1%, respectively (Table 1). Only a few among the 1228 proteins identified in the uninduced and the 1231 in the induced condition were selectively expressed (Supplemental Fig. S4); these included several members of the VirB/D4 T4SS in the induced condition. Compared to the expressed transcriptome, the proteome coverage reaches 90% for both the uninduced and induced condition.

Although each experimental and computational approach contributed unique protein identifications to the final data set (see Supplemental Fig. S5), for similar studies aiming to maximize coverage of an expressed proteome with a minimum number of experiments, we recommend use of subcellular fractionation (Cyt and TM), and performing OGEPep and measuring each fraction twice using the exclusion list approach. This approach would identify 1153 proteins, i.e., 92%, while requiring only 15% of the mass spectrometry runs needed to identify all 1250 proteins.

Evidence for having reached an expressed proteome endpoint Several lines of evidence indicated that the 1250 distinct protein groups are very close to the complete proteome endpoint that is actively expressed under the investigated conditions.

First, a comparison of the total number of PSM identifications showed that MS-GF+ added 67% more PSMs than Mascot-Percolator (Supplemental Fig. S3A). Yet, at the level of distinct peptides, this increase was smaller (+37%) (Supplemental Fig. S3B) and amounted to a mere 3%, or 33 additional proteins at the protein level (Supplemental Fig. S3C), despite having added several hundred thousand additional PSMs. Using a third search engine, Sequest, would have only added one additional protein for all experimental spectra. This indicates that, similar to the transcriptome, we have also measured the expressed proteome to saturation. The exponential model fitted to the eight OGEPep experiments (Fig. 3A) supports this: Doubling the number of PSMs on OGEPep samples (roughly 305,000 additional PSMs, i.e., ~36% more PSMs overall) would only identify five new proteins (red number on top of red dashed line, Fig. 3A).

Second, our expressed proteome encompassed all proteins identified in three previous B. henselae proteomics studies (Rhomberg et al. 2004; Eberhardt et al. 2009; Li et al. 2011), while adding many more low-abundance proteins (Supplemental Fig. S6A–C).

Third and most importantly, a comparison of the protein parameter distributions of the data sets expressed protein-coding genes (1353) and final expressed proteome (1250) showed that there is virtually no underrepresentation anymore in those areas of the proteome that we had specifically targeted; i.e., ADE successfully eliminated these differences present in the OGEPep pilot study (Supplemental Fig. S7). Two examples illustrate this point: (1) For the parameter isoelectric point (pl), basic proteins are underrepresented in the OGEPep data set. After carrying out the ADE approach, there is only a small difference between the densities of the data set “final” and “expressed” (Supplemental Fig. S7, top panels); and (2) for the parameter gravy, membrane proteins with one or more predicted transmembrane domains (gravy values above 0.5) are underrepresented in the OGEPep data set. Again, after the ADE approach, the densities for the data sets “expressed” and “final” are virtually identical (Supplemental Fig. S7, middle panels). This comparison also showed that ADE could add proteins encoded by genes that are expressed at lower levels under the conditions studied (Supplemental Fig. S7, last panels). Two-dimensional density plots of the gene expression level versus the parameters length, pI, and gravy (Supplemental Fig. S8) for the data set final expressed proteome (1250) versus not seen proteins (217) showed that there is still a noticeable tendency for short and basic proteins to be enriched among genes with expression levels close to the threshold whose proteins were not identified (Supplemental Fig. S8A,B). These are not expected to be detectable with the shotgun proteomics approach since short and basic proteins have fewer tryptic peptides in the detectable range of the mass spectrometer. In contrast, for the two-dimensional density plot with the protein parameter gravy (values above 0.5 are found in proteins with transmembrane domains), we observed no bias (Supplemental Fig. S8C), indicative of a complete membrane proteome coverage.

Table 1. Summary of identified PSMs, peptides, and proteins and estimated FDR levels

<table>
<thead>
<tr>
<th>Class</th>
<th>No. of PSMs</th>
<th>No. of distinct peptides</th>
<th>No. of distinct proteins*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>747,352</td>
<td>43,193</td>
<td>1240</td>
</tr>
<tr>
<td>3a</td>
<td>7356</td>
<td>283</td>
<td>10</td>
</tr>
<tr>
<td>3b</td>
<td>12,161</td>
<td>663</td>
<td>n.a.</td>
</tr>
<tr>
<td>Total B. henselae</td>
<td>766,869</td>
<td>44,139</td>
<td>1250</td>
</tr>
<tr>
<td>Decoy hits</td>
<td>54</td>
<td>42</td>
<td>7</td>
</tr>
</tbody>
</table>

*Protein groups identified by 3a peptides are unique protein sequences that can be encoded by two or more distinct gene models. Not those implied by ambiguous class 3b peptides (n.a.).

The total number of PSMs, distinct peptides, and distinct proteins is shown, further separated by peptide evidence class (Grobei et al. 2009). We only considered proteins implied by class 1a and 3a peptides, not those implied by ambiguous class 3b peptides (n.a.).
Results - Research article II

Covering complete expressed prokaryotic proteomes

To extend the evolutionary conservation analysis beyond members of the genus *Bartonella*, we relied on the eggNOG resource, which contains orthology information from 1133 organisms, including *B. henselae* (Powell et al. 2012). Among the 1488 *B. henselae* proteins, only 55 proteins lack any functional annotation; they are a subset of the 395 without ortholog (black bars, third ring, Fig. 5). Strikingly, 52 of these 55 were not detected, again a significant enrichment (*P*-value < 10^{-3}). A significant number of the genes (16) encoding these 55 proteins clustered in a region from 1612–1674 kbp that harbors 59 predicted ORFs (*P*-value < 10^{-3}) (yellow box, Fig. 5). Location in this plastic, repeat-rich genome region (orange bars, fourth ring) may lead to strong transcription of genes that do not represent a bona fide protein-coding ORF.

The evolutionary conservation information provided by eggNOG, together with high-quality experimental proteomics data, represents a particular useful combination to identify candidates for overpredicted protein-coding genes in genome annotations: The densities of the protein length distribution of the proteins not identified (217) were clearly separated from that of the proteins seen (1250) (Supplemental Fig. S9A). Among the proteins not seen, those that lack any functional annotation are considerably shorter than those with a functional annotation (Supplemental Fig. S9B). Since we can detect short proteins with our set-up (see density of the 150 shortest proteins detected compared to all, Supplemental Fig. S9C), the proteins that lack an ortholog and any functional annotation may either only be expressed under different conditions or are potential overpredicted ORFs.

Coverage of the membrane proteome and the VirB/D4 T4SS

The membrane proteome serves many essential roles in cellular communication, transport, adhesion to host cells, and evasion of the host immune system. While accounting for up to one third of the gene products, >50% of the druggable targets fall into this category (Hopkins and Groom 2002). However, due to the amphipathic nature and low abundance of membrane proteins, they are notoriously underrepresented in proteomics studies (Poetsch and Wolters 2008; Tan et al. 2008; Hellwig et al. 2010).

To reach a high protein sequence coverage for membrane proteins, we used a combination of trypsin and chymotrypsin in all membrane samples and, furthermore, applied proteolytic digestion in 60% (v/v) methanol to improve cleavability of hydrophobic proteins (Fischer et al. 2006; Supplemental Methods). Among 924 proteins identified in the first four pilot phase experiments (63% of all distinct proteins), 182 contained predicted transmembrane domains (54%) (Fig. 6A, left panel). However, the ADE approach was able to eliminate this underrepresentation of membrane proteins: among the final 1250 identified proteins (85% of all distinct annotated proteins), 289 of the 338 distinct proteins with one or more predicted transmembrane regions were found, i.e., 86% (Fig. 6A, right panel; Supplemental Fig. S10A). Notably, the OGEpep fractionation regimen was particularly successful in identifying membrane proteins. We also identified 54 of the 58 predicted secreted proteins (95%). These include many proteins for which PSORTb (Yu et al. 2010) predicts localization in the membrane space and where other studies could confirm their localization in inner or outer membrane, periplasm, or the extracellular space (Supplemental Fig. S10B). Together with the striking result that transmembrane proteins with high gravy values are not overrepresented among the 217 nonidentified proteins compared to 1250 seen proteins (see Supplemental Fig. S8C), the data sug-

![Figure 4. Correlation of gene expression strength and successful protein identification rate.](image-url)
Results

Among these 125, 36 transmembrane and 12 secreted proteins were found, a significant enrichment (P-value < 0.0018) compared to 343 membrane and secreted proteins among the 1250 proteins. A striking feature was the strong regulation of different families of autotransporters, which rely on the type V secretion pathway for their delivery to the surface of Gram-negative bacteria (Leyton et al. 2012). These included two representatives of prokaryotes has important implications for studying their expression under different conditions in a quantitative fashion. Ideally, such a task would be performed with the more sensitive targeted proteomics approach (Schmidt et al. 2011), which typically relies on predicted proteotypic peptides (PTPs) using tools like PeptideSieve (Mallick et al. 2007). Our data indicate that a comprehensive discovery proteomics approach adds clear value with respect to experimentally identified PTPs, as we could identify peptides for 145 proteins for which PeptideSieve predicted no PTP (see Supplemental Methods). We provide the proteomics and transcriptomics data with results of several prediction algorithms (Supplemental Table S5A), and all experimentally identified peptides (Supplemental Table S5B), from which the best-suited PTPs can be selected using available guidelines (Picotti and Aebersold 2012).

Identification of differentially expressed proteins

Our in-depth proteome analysis precluded the measurement of biological replicates. We thus relied on DESeq to identify the most significantly differentially regulated proteins between induced and uninduced states (see Methods). The top 10% differentially expressed proteins (Supplemental Table S6), including 68 up-regulated (red dots), and 57 down-regulated proteins (green dots) in the induced condition, are highlighted in Figure 7.

When we assessed the level of induction at the RNA and protein level, we observed that the induction of virB/D4 and bep operons, which are direct targets of the transcriptional regulator BatR, seemed to be more prominent at the protein level. They also included more cases with statistical significance of the up-regulation (Fig. 6B, log2 fold changes, left panel). A comparison of the log2 fold changes at the RNA level versus those at the protein level indicated that several of the virB/D4 and bep genes appear to be regulated preferentially at the post-transcriptional level, indicated in Figure 6C by their position close to the vertical axis.

The ability to identify complete membrane proteomes of prokaryotes has important implications for studying their expression under different conditions in a quantitative fashion. Ideally, such a task would be performed with the more sensitive targeted proteomics approach (Schmidt et al. 2011), which typically relies on predicted proteotypic peptides (PTPs) using tools like PeptideSieve (Mallick et al. 2007). Our data indicate that a comprehensive discovery proteomics approach adds clear value with respect to experimentally identified PTPs, as we could identify peptides for 145 proteins for which PeptideSieve predicted no PTP (see Supplemental Methods). We provide the proteomics and transcriptomics data with results of several prediction algorithms (Supplemental Table S5A), and all experimentally identified peptides (Supplemental Table S5B), from which the best-suited PTPs can be selected using available guidelines (Picotti and Aebersold 2012).

Identification of differentially expressed proteins

Our in-depth proteome analysis precluded the measurement of biological replicates. We thus relied on DESeq to identify the most significantly differentially regulated proteins between induced and uninduced states (see Methods). The top 10% differentially expressed proteins (Supplemental Table S6), including 68 up-regulated (red dots), and 57 down-regulated proteins (green dots) in the induced condition, are highlighted in Figure 7.

Among these 125, 36 transmembrane and 12 secreted proteins were found, a significant enrichment (P-value < 0.0018) compared to 343 membrane and secreted proteins among the 1250 proteins. A striking feature was the strong regulation of different families of autotransporters, which rely on the type V secretion pathway for their delivery to the surface of Gram-negative bacteria (Leyton et al. 2012). These included two representatives of prokaryotes has important implications for studying their expression under different conditions in a quantitative fashion. Ideally, such a task would be performed with the more sensitive targeted proteomics approach (Schmidt et al. 2011), which typically relies on predicted proteotypic peptides (PTPs) using tools like PeptideSieve (Mallick et al. 2007). Our data indicate that a comprehensive discovery proteomics approach adds clear value with respect to experimentally identified PTPs, as we could identify peptides for 145 proteins for which PeptideSieve predicted no PTP (see Supplemental Methods). We provide the proteomics and transcriptomics data with results of several prediction algorithms (Supplemental Table S5A), and all experimentally identified peptides (Supplemental Table S5B), from which the best-suited PTPs can be selected using available guidelines (Picotti and Aebersold 2012).

Identification of differentially expressed proteins

Our in-depth proteome analysis precluded the measurement of biological replicates. We thus relied on DESeq to identify the most significantly differentially regulated proteins between induced and uninduced states (see Methods). The top 10% differentially expressed proteins (Supplemental Table S6), including 68 up-regulated (red dots), and 57 down-regulated proteins (green dots) in the induced condition, are highlighted in Figure 7.

Figure 5. Integration of expression evidence with structural genome information and evolutionary conservation. Genes whose proteins were not identified cluster in specific regions of the B. henselae genome. Outer ring: Genes whose proteins were identified (light blue) or not identified (red). Second ring: Protein-coding genes classified by the RNA-seq analysis as expressed (gray) or not (dark green). Third ring: Genes without a detectable ortholog among species of lineage 4 of the genus Bartonella (Engel et al. 2011) (turquoise) and genes without any functional annotation by the eggNOG classification (black). Fourth ring: Repeat regions identified by RepSeek (orange) (Vallenet et al. 2006) and rRNA repeat regions (light orange). Fifth ring: Location of a prophage region (ochre), three genomic islands (blue), the virB/D4 and trw operons (sky blue), and a novel genomic region enriched in repeats as well as highly expressed genes whose encoded proteins were not identified (yellow). The results of hypergeometric tests for selected data sets are also shown (asterisks indicate statistically significant enrichment; see text). For the hypergeometric test, we used all possible protein-coding genes for the identified “seen” proteins (1250 distinct proteins encoded by 1261 gene models) and “not seen” proteins (217 distinct proteins encoded by 227 gene models). The circular plot was generated using DNAPlotter (Carver et al. 2009).
of the trimeric autotransporter adhesins (BH01490, BH01510), a class of virulence factors essential for *Bartonella* pathogenicity (Franz and Kempf 2011). Furthermore, seven of 10 proteins with an autotransporter beta domain (as predicted by SMART version 7) (Letunic et al. 2012) were among the top 10% differentially regulated proteins (six up-regulated, one down-regulated) (yellow dots, Fig. 7; Supplemental Table S6), i.e., a significant enrichment (P-value < 4 x 10^{-7}). BH13020, BH13180, and BH13010 were the top three up-regulated proteins, which ranked even higher than members of the *virB/D4* operon. While less is known about the role of this family of autotransporters in *Bartonella*, they were found to be up-regulated during infection of endothelial cells (Quebatte et al. 2010) and may be involved in adhesion to host cells (Litwin et al. 2007). Finally, two of the four outer membrane proteins of the hemin binding protein family (HbpC and HbpB) were found. HbpC was shown to protect *B. henselae* against hemin toxicity and to play a role during host infection (Rodden et al. 2012).

The top 10% regulated proteins included six of the seven Beps and all *VirB/D4* T4SS proteins except *VirB3*. For this small protein (103 amino acids) with one predicted transmembrane domain, we only found four spectra, all in the induced condition. This indicates that a large experimental effort is required to detect proteins that combine several parameters which complicate their mass spectrometric identification with shotgun proteomics, i.e., they are short, basic, and hydrophobic. Another protein exclusively identified in the induced condition is BH13250, a hypothetical protein with a transmembrane domain (Supplemental Table S6). Its location just upstream of the *virB/D4* operon is conserved in other *Bartonella*, suggesting that it may potentially carry out a yet to be determined function as a virulence factor. Finally, another interesting up-regulated protein is RpoH1 (BH15210), an alternative RNA-polymerase sigma factor 32. A role in virulence has been documented for its gene in an in vivo mouse infection model for the closely related *Brucella* (Delory et al. 2006).

Figure 6. Membrane proteome coverage and dynamics. (A) Comparison of the membrane proteome coverage achieved in four pilot experiments (left panel) and the final data set (right panel). Membrane proteins are binned according to the number of predicted transmembrane domains; the percentage of proteins identified per bin is shown above each bar. The legends summarize the respective coverage achieved comparing the respective data set (pilot phase/final) against all distinct proteins and for the subset of proteins with transmembrane domains. Membrane proteins are underrepresented in the pilot phase but not in the final data set. (B) Transcript and protein expression changes of the *virB/D4* T4SS and downstream *bep* operon. Operon structures (upper panel) are drawn to scale. The lower left panel shows the log2 fold changes at the transcript and protein level for the induced versus uninduced state (the ∞ indicates that the protein was only identified in the induced condition). Fold changes and significance were calculated with DESeq. Regulation at the protein level appears to be more pronounced compared to the transcript level. The lower right panel visualizes the protein expression changes upon induction onto a schematic representation of the assembled *VirB/D4* T4SS using different shades of blue. (C) Comparison of expression changes at transcript and protein level. The respective log2 fold changes based on the RPKM values and normalized spectral counts are shown. Members of the *VirB/D4* T4SS are shown in blue (BH13360 in light blue), *Bartonella* effector proteins (Beps) in dark blue. Three proteins that exhibited the most significant differential expression (Supplemental Table S6; Fig. 7) are also shown with their identifiers.
Results

Using a discovery proteomics approach, the expressed proteome of *B. henselae* was exhaustively studied under two conditions that mimic those encountered in different hosts. The saturated transcriptome analysis of RNA extracted from matched samples provided the best possible endpoint estimate for the number of actively transcribed protein-coding genes. ADE was able to virtually eliminate the biases of commonly underrepresented short, basic, and particularly lower-abundance and membrane protein classes, all of which are experimentally tractable. Based on a very stringent FDR at the PSM level, we identified 85% of all distinct, annotated proteins, and ~90% compared to the expressed protein-coding genes in the two conditions. Several lines of evidence indicated that this is very close to all proteins that can be identified by a discovery proteomics approach with current technology. This is best illustrated by the complete membrane proteome coverage, including evidence for all members of the important VirB/D4 T4SS. The analysis of the genome organization revealed that genes whose transcripts were detected, but not their corresponding protein products, were highly enriched in genomic islands. Information regarding evolutionary conservation provided evidence for preferential expression of genes with a predicted ortholog. In contrast, genes that lacked an ortholog and functional annotation were mostly not observed at the protein level, suggesting possible overprediction in genome annotations.

**Conclusion**

Our report is the second complete expressed proteome reported (de Godoy et al. 2008). Using a similarly extensive fractionation strategy, our matched transcriptomics and proteomics data correlated quite well (r = 0.57) while identifying the VirB/D4 T4SS as a prominent target of post-transcriptional regulation. The rigorous approach to sequence transcriptome and proteome to saturation and to provide proof for having eliminated observed biases at the protein level is unique. It supports a recent perspective article showing that up to 90% of an expressed proteome (“nearly complete”) can be measured quite quickly (Mann et al. 2013), but the remaining 10% require extensive effort. It also underscores that the difference between “comprehensive” and “complete” can be quite large, in particular with respect to coverage of the membrane proteome (Beck et al. 2011). The higher coverage of distinct annotated proteins (85%) compared to the proteome expressed by haploid and diploid yeast (67%) suggests that prokaryotes express a higher fraction of the encoded proteins, potentially reflecting their need to quickly adapt to changing conditions. This fraction may be lower for more complex prokaryotes.

The data attest to the value of a discovery proteomics approach in providing experimentally identified PTPs beyond those predicted in silico. The sensitive quantitative measurement of such PTPs by SRM holds particular promise to be able to screen entire bacterial surfacesomes and to identify targets for novel anti-infectives. Ideally, such studies would be carried out using in vivo infection models. Enabled by the consideration of organism-specific peptide information (Delmotte et al. 2010), they will bring the analysis of mixed in vivo proteomes within reach and complement the power of dual RNA-seq (Westermann et al. 2012) for this task. We expect that the strategy described here will be useful for some of these exciting applications.

**Methods**

**Bacterial growth and subcellular fractionation**

The *B. henselae* strain MQB307 harbors a deletion of the response regulator *batR* (BH00620) and its cognate sensor histidine kinase *batS* (BH00610) and carries a plasmid-encoded copy of *batR* under the control of an IPTG-inducible promoter (for details, see Supplemental Methods; Supplemental Tables S1, S2). MQB307 was grown on Columbia blood agar (CBA) plates supplemented with 30 mg/L kanamycin with (induced condition) or without (uninduced condition) 500 μM IPTG at 35°C and 5% CO₂ for 60 h. The subcellular fractionation was performed as previously described (Rhomberg et al. 2004; Supplemental Methods). To maximize the recovery of membrane proteins, the total membrane fraction (TM) was further separated into inner membrane (IM) and outer membrane (OM) fraction.
RNA extraction and whole transcriptome sequencing

RNA was isolated from bacterial cells as described (Quebatte et al. 2010). Whole transcriptome libraries were produced using the RiboMinus Bacterial Transcriptome Isolation Kit (Life Technologies), and the SOLiD Total RNA-seq kit (Applied Biosystems). Briefly, cDNA libraries were size-selected and amplified for 18 cycles of PCR. The whole transcriptome library was used for emulsion-PCR based on a concentration of 0.5 pM. Sequencing beads were pooled and loaded on a full SOLiD-4 slide; between 55–87 million 50-base sequencing reads were generated per library (Supplemental Table S3). For details, see Supplemental Methods.

RNA-seq data processing and transcriptome coverage analysis

The sequenced reads were mapped to the genome sequence of the B. henselae Houston-1 strain using the BioScope 1.3.1 mapping pipeline. Among all uniquely mapping reads, those of lower quality were removed (for more detail, see Supplemental Methods; Supplemental Fig. S12). The count data summary for annotated B. henselae ORFs was generated using the HTSeq package. To create Figure 2A, the filtered reads were shuffled and sequentially mapped to the genome; a protein-coding ORF was classified as expressed when accumulating five or more distinct reads in the 5’ end of the ORF. Based on this data, nonlinear regression models were constructed to estimate the effect of doubling the number of reads. For details, see Supplemental Methods.

Protein and peptide fractionation and mass spectrometry

The subcellular fractions (Cytoplasm, Transmembrane, IMenvelope, OMenvelope) were further fractionated biochemically, including OFFGEL electrophoresis at the protein (OGEprot) and peptide level (OGEpep), and size exclusion chromatography (SEC, “gel filtration”). To enrich for low-abundance proteins, we used the ProteoMiner approach (Guerrier et al. 2008). More detail on the biochemical fractionations, digest conditions, and the mass spectrometry set-up is given in the Supplemental Methods and in Supplemental Figure S13. Samples were injected into a NanoLC HPLC system (Eksigent Technologies) by an autosampler, separated on a self-made reverse-phase tip column packed with C18 material, and acquired on an LTQ Orbitrap XL or LTQ FT Ultra mass spectrometer (both Thermo Scientific).

Database searching and data processing

To minimize the chance for false positive assignments, spectra were searched against a combined database (1488 B. henselae proteins, 3336 sheep proteins, a positive control [myc-gfp], and sequences of 256 common contaminants [keratins, trypsin, etc.]) either with Mascot (version 2.3.0, Matrix Science) or with MS-GF+ (MS-GFDB v7747). For Mascot, data were further post-processed with Percolator (Brosch et al. 2009). Based on the target-decoy search approach, a Percolator/MS-GF+ score cutoff was determined that resulted in an estimated 0.01% FDR at the PSM level. All PSMs above this cutoff were classified with the PeptideClassifier software (Qeli and Ahrens 2010), and only peptides (tryptic or semitryptic) that unambiguously imply one bacterial protein sequence were considered (Table 1). For details, see Supplemental Methods.

ADE analysis

Exponential curves were fitted to each block of experiments with a shared biochemical fractionation regimen to find a saturation threshold (Fig. 3A). We then used this fit to predict the saturation beyond the point of experimentally observed PSMs for each biochemical fractionation regimen (Fig. 3A, dashed lines). For details on the exponential model, approximating confidence bands, density estimation of physicochemical parameters, and computation of physicochemical parameters and other protein sequence features, see Supplemental Methods.

Statistical analysis

Statistical tests were performed using the statistical software R 2.15.2 (www.R-project.org). All reported P-values are from geometric tests and are adjusted for multiple testing controlling the corresponding FDR (Benjamini and Hochberg 1995). Significance is based on an alpha level of 5%.

Transcript and protein abundance estimation

Transcript abundance was estimated via RPKM values calculated similar to Mortazavi et al. (2008). The sum of mapped and filtered reads per gene was divided by its length (in kilobases) and the sum of reads for all B. henselae protein-coding genes (in million reads). Relative protein abundance (in ppm) (see Supplemental Fig. S6C) was estimated based on spectral counts as described (Schrimpf et al. 2009).

Orthologs, sequence repeats, and functional protein classification

Orthologous genes conserved in B. henselae, B. tribocorum, and B. grahamii were taken from Engel et al. (2011). To find duplicated regions of 50 nt or longer in the B. henselae genome, we used RepSeek (version 6.5) (Achaz et al. 2007). For functional protein classification, we relied on the eggnOG resource (http://eggnog.embl.de). For details, see Supplemental Methods.

Differential expression analysis

Differential transcript and protein expression analysis was carried out with the R package DESeq (version 1.6.1) (Anders and Huber 2010). Our description of condition-specific complete expressed proteomes precluded the analysis of biological replicates. Since DESeq ranks proteins according to statistical significance, i.e., the top-ranked proteins are observed by many spectra, we minimized the potential to erroneously identify differentially expressed proteins by chance. On the other hand, without replicates, we lack the power to detect lower expressed, truly differentially regulated proteins.

Data access

RNA-seq data have been submitted to the NCBI Genome Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under the GEO Series accession number GSE44564. Proteomics data associated with this manuscript can be downloaded from ProteomeXchange (http://proteomcentral.proteomeexchange.org/) under accession number PXD000153.

Acknowledgments

We thank Dr. Ermir Qeli for initial work on the project, Sisi Wu and Dr. Liljana Pasa-Tolic (PNNL, USA) for helpful discussions, Dr. Sangtae Kim (PNNL, USA) for the MS-GF+ software, and Drs. Bernd Wollscheid (IMSB, ETH Zürich), Aurelien Carlier, and
Gabriella Pessi (Institute of Plant Biology, UZH) for critical reading of the manuscript. C.H.A. acknowledges support from the Swiss National Science Foundation (SNSF) under grant 31003A_130723. C.D. acknowledges support from the SNSF under grant 31003A_132979 and from SystemsX.ch under grant 5IRT_0_126008 (RTD InfectX). A part of the research was performed using EMSL, a national scientific user facility sponsored by the Department of Energy's Office of Biological and Environmental Research and located at Pacific Northwest National Laboratory.

References


Received November 18, 2012; accepted in revised form July 17, 2013.
3.3. RESEARCH ARTICLE III (published)

Dual input control: Activation of Bartonella henselae VirB/VirD4 Type IV secretion system by the stringent sigma factor RpoH1 and the BatR/BatS two component system

Maxime Quebatte¹, Mathias S. Dick¹, Volkhard Kaever, Alexander Schmidt and Christoph Dehio

¹ These authors contributed equally

Molecular Microbiology 2013 Nov; 90(4):756-75

3.3.1. Statement of the own participation

I designed all experiments presented in the research article III, many of which have been performed by Mathias Dick, a master student under my supervision. I designed and performed the initial screening and mapping of the transposon mutants (Table S1), contributed to the preparation of cellular extracts for ppGpp concentration determination (Fig. 2) and to the infections and quantification of invasome formation (Fig. 6 and 10) together with Mathias Dick. I performed the sample preparation for Mass Spectrometry, which were processed by Alexander Schmidt from the Proteomics Core Facility of the Biozentrum, who also provided the primary analysis for the label-free quantification of the detected proteins. I performed the analysis of the differentially regulated protein as well as the overall data analysis and the validation of the Trw T4SS expression data (Fig. 11). I wrote the manuscript and assembled all the figures presented in this article.
Dual input control: activation of the Bartonella henselae VirB/D4 type IV secretion system by the stringent sigma factor RpoH1 and the BatR/BatS two-component system

Maxime Québatte, Mathias S. Dick, Volkhard Kaever, Alexander Schmidt and Christoph Dehio

Introduction

The co-ordinated expression of virulence factors during the progression of the infection cycle is an essential aspect of bacterial pathogenicity and a prerequisite for successful host-adaptation (Gross, 1993; Cotter and DiRita, 2000; Beier and Gross, 2006; Boutte and Crosson, 2013). The stringent response (SR) represents a prominent adaptive response mechanism that has been recruited by bacterial pathogens to respond to the microenvironments encountered in the host and to control expression of their virulence factors (Dalebroux et al., 2010). It relies on the cytosolic accumulation of the second messengers guanosine tetraphosphate [both referred here as (p)ppGpp] produced by enzymes of the RelA/SpoT homologue (RSH) superfamily (Mittenhuber, 2001; Atkinson et al., 2011). Many of the effects mediated by the SR are the consequence of the direct interaction between the core RNA polymerase (RNAP) and (p)ppGpp (Ross et al., 2011; Zuo et al., 2013) which can result in either transcription activation or deactivation at specific promoters (Artisimovitch et al., 2004; Lemke et al., 2011). In Gram-negative bacteria, those transcriptional alterations can be amplified by DksA, a small protein that binds to the secondary channel of the RNAP where it is thought to interact with (p)ppGpp and to stabilize its interaction with the polymerase (Paul et al., 2004; Perederina et al., 2004).

The use of alternative sigma factor(s) represents another common strategy to control the expression of virulence factors in response to environmental changes (Kazmierczak et al., 2005; Rowley et al., 2006; Dong and Schellhorn, 2010; Österberg et al., 2011; Shingler, 2011). It relies on the ability of these alternative factors to confer a different range of promoter selectivity to the RNAP compared with the housekeeping σ70 (Österberg et al., 2011). The expression and availability of alternative sigma factors is usually regulated by complex regulatory networks that can integrate several signals (Österberg et al., 2011; Ho and Ellermeier, 2012).

Bacteria of the α-proteobacterial genus Bartonella are haemotropic facultative intracellular pathogens that share a common infection strategy. Typically, Bartonella spp. are transmitted to their restricted mammalian reservoir host...
by the contaminated faeces of a blood-sucking arthropod followed by an intradermal inoculation of the released bacteria (Chomel et al., 2009). The bacteria first colonize an elusive primary niche, a part of which was proposed to be constituted by endothelial cells (Dehio, 2005). Subsequently, bartonellae are seeded into the bloodstream where they bind to and invade red blood cells resulting in long-lasting intraerythrocytic infections (Seubert et al., 2002; Harms and Dehio, 2012). The success of this infection cycle relies on the successive adaptation to the different niches encountered together with stage-specific expression of dedicated virulence factors. These comprise the VirB/D4 type IV secretion system (T4SS) and its secreted substrates (Bartonella effector proteins or ‘Beps’) (Schulein and Dehio, 2002; Schmid et al., 2004), the Trw T4SS (Seubert et al., 2003; Vayssier-Taussat et al., 2010), the trimeric autotransporter BatA (Zhang et al., 2004; Müller et al., 2011) and other type V secretion systems (T5SS) (Seubert et al., 2002; Litwin et al., 2007; Saenz et al., 2007; Guy et al., 2013), or members of the haemin-binding protein family (Hbp) (Battisti et al., 2006; Roden et al., 2012). To date, the underlying mechanisms co-ordinating the expression of Bartonella henselae virulence factor in response to host interaction remain largely elusive. Bartonella Hbps have been shown to undergo differential regulation in response to temperature and haem concentration (Battisti et al., 2006; Roden et al., 2012) by a process likely regulated by the iron responsive regulators RirA and Irr (Battisti et al., 2007) that could allow the distinction between the arthropod vector and the mammalian host environment. Previously, we have shown that the BatR/BatS TCS directly controls the pH-restricted expression of the VirB/D4 T4SS and its secreted substrates in B. henselae (Quebatte et al., 2010), leading to the model that physiological pH (pH 7.4) could represent a specific activation signal for the bacteria in their mammalian host, in contrast to the alkaline environment encountered in the arthropod midgut. Yet, this model could not explain how bacteria would distinguish early and late stages of host infection. At an early stage, which includes invasion of endothelial cells, the VirB/D4 T4SS and its secreted substrates are required (Schulein and Dehio, 2002; Saenz et al., 2007; Vayssier-Taussat et al., 2010) whereas they are dispensable after the release of the bacteria into the bloodstream (Schulein and Dehio, 2002). In contrast, the Trw T4SS is necessary for the establishment of the intraerythrocytic niche but dispensable for the early stages of host colonization (Seubert et al., 2003; Vayssier-Taussat et al., 2010). Moreover, the slow response curve observed for the BatR/BatS regulated genes in response to pH shift remained unexplained.

Here we used the zoonotic pathogen B. henselae as a model to decipher the mechanisms of niche adaptation in the context of host–pathogen interaction, with a special focus on the regulation of the VirB/D4 T4SS and its secreted substrates during endothelial cell infection. By performing a genetic screen in conditions where the BatR regulon was induced in a pH-dependent manner in the absence of host cells (host cell-free induction) (Quebatte et al., 2010), we identified the SR components SpoT and DksA as key regulators of B. henselae virulence. We further showed that the SR contribution to the VirB/D4 T4SS regulation is mediated by the alternative sigma factor RpoH1. Epistasis analysis demonstrated that the BatR/BatS TCS and the alternative sigma factor RpoH1 form a dual input switch where both components are essential but neither is sufficient for the induction of the VirB/D4 T4SS. Finally, using a shotgun proteomics approach we showed that the RpoH1 regulon represents a subset of the SpoT/DksA regulated proteins and that the VirB/D4 T4SS and its secreted substrates are among the few proteins that require the dual activation of the BatR/BatS TCS and RpoH1 for their expression. Together, our data demonstrate that a dual signal input is necessary for the induction of the B. henselae VirB/D4 T4SS and offer new insights into the strategy used by this pathogen to recognize its host environment.

Results

A genetic screen identifies the SR components spoT and dksA as key components for the induction of the B. henselae VirB/D4 T4SS

Our previous characterization of B. henselae virB promoter regulation using a transcriptional fusion to gfp (PvirB::gfp) combined with flow cytometry allowed us to establish host cell independent conditions to monitor the induction of this promoter (Quebatte et al., 2010). Aiming at identifying putative factors involved in the regulation of the B. henselae virB operon additionally to the known BatR/BatS TCS, we generated a tester strain harbouring a chromosomal copy of the PvirB::gfp reporter in the B. henselae wild-type background and subjected the resulting strain to mutagenesis using a Himar1 transposon (Saenz et al., 2007). Individual mutants were assessed for loss or attenuation of the reporter expression after 48 h of host cell-free induction by flow cytometry (see Experimental procedures). As already described for the plasmid-encoded reporter (Quebatte et al., 2010) we observed the presence of two populations in wild-type bacteria: a major population expressing high levels of GFP (‘on-state’) and a smaller, yet variable fraction of bacteria with very low GFP expression levels (‘off-state’). In contrast, the induction profile of PvirB::gfp in the background of a ΔbatR mutant resulted in a single population of bacteria with an intermediate expression level lying between the on- and the off-state of wild-type bacteria (Fig. S1). Whereas most transposon mutants...
behaved like wild-type with respect to \( P_{\text{virB}} \) induction, two major alterations could be observed: class I mutations showed a shift in the on/off ratio and class II mutations resulted in the establishment of an intermediate GFP expression level, as described for a \( \Delta \text{batR} \) mutant. We selected all class I mutants showing less than 30% of GFP positive bacteria and all class II mutants for further characterization. From a total of 1920 transposon mutants assessed, 56 displayed a reproducible defect in \( P_{\text{virB}} \) induction (Table S1). The sites of transposon insertion were mapped by arbitrary PCR and subsequent sequencing for 48 mutants, resulting in the identification of 36 different loci (Table S1), including three insertions in the chromosomal \( P_{\text{virB}}:gfp \) reporter and two insertions within \( \text{batS} \), the histidine kinase of the BatR/BatS TCS. The latter represented the only two class II mutants isolated in this screen. Grouping of the mutants into functional categories based on the predicted function of the disrupted gene suggested that various pathways contribute to the regulation of \( P_{\text{virB}} \) induction (Table S1). In our screen we isolated four independent mutants related to the secondary messengers (p)pGpp. Three insertions disrupted \( \text{spoT} \) (BH05040), encoding the sole representative of the RelA/SpoT homologue (RSH) superfamily (Mittenhuber, 2001; Atkinson \textit{et al}., 2011) in \( B. \text{henselae} \). A fourth insertion affected the putative promoter region of \( \text{dksA} \) (BH09660), encoding a homologue of the DnaK suppressor protein A, which is generally considered to be a cofactor of (p)pGpp (Perederina \textit{et al}., 2004). Based on these findings, we hypothesized that mounting of the so-called SR would constitute a key signal for the induction of \( B. \text{henselae} \) VirB/D4 T4SS and focused on this pathway for further investigations.

\( B. \text{henselae SpoT and DksA are both required for the full induction of the VirB/D4 T4SS} \)

To confirm the involvement of SpoT and DksA in the regulation of \( B. \text{henselae} \) VirB/D4 T4SS, scarless in-frame deletion mutants of \( \text{spoT} \) and \( \text{dksA} \) were generated in the background of the reporter strain. Flow cytometry analysis confirmed that the deletion strains phenocopied the isolated transposon mutants with a marked decrease in \( P_{\text{virB}} \) expression for both mutants compared with the wild-type level (Fig. 1A). Single cell data analysis indicated a less drastic effect for the \( \text{spoT} \) deletion compared with the \( \text{dksA} \) deletion, with a small proportion of the \( \Delta \text{spoT} \) mutant bacteria still showing a higher GFP fluorescence (Fig. 1B). The defect in either mutant was largely complemented by supplementing a plasmid-encoded copy of the deleted gene under the control of \( P_{\text{taclac}} \) (\( \text{pspoT} \) and \( \text{pdk} \)). In both cases complementation was most efficient without addition of IPTG, suggesting that leaky plasmid expression in the

---

Fig. 1. \( B. \text{henselae spoT and dksA are required for} \ P_{\text{virB}} \) activation.

A. Expression of the chromosomally encoded \( P_{\text{virB}}:gfp \) reporter after 48 h of induction in M199/10% FCS. GFP fluorescence values were determined by flow cytometry for wild-type, \( \Delta \text{spoT} \) or \( \Delta \text{dksA} \) bacteria carrying no plasmid (1), empty plasmid (2), \( \text{pspoT} \) (3) or \( \text{pdk} \) (4) with addition of 500 \( \mu \text{M} \) IPTG (black bars) or without (white bars). Values represent the mean ± SD of the GFP fluorescence values (geometric mean) from three independent experiments expressed as per cent of wild-type GFP fluorescence intensity (without plasmid). Asterisks indicate statistically significant difference to wild-type bacteria without plasmid (\( P < 0.01 \)) as determined by one-way ANOVA and Tukey’s post hoc tests.

B. Representative flow cytometry data for the samples presented in (A) and displaying the side light scatter (SSC-H) and the GFP intensity (FL1-GFP) as double logarithmic dot-blot. The dashed line is set for comparison purpose.

---

© 2013 John Wiley & Sons Ltd, \textit{Molecular Microbiology}, \textbf{90}, 756–775
absence of the inducer was sufficient to restore function-
ality whereas higher expression levels could interfere
with the $P_{\text{virB}}$ activity (Fig. 1A). To assess a possible hier-
archical relationship between spoT and dksA, we also
 evaluated the effects of DksA overexpression in $\Delta$spoT
and SpoT overexpression in $\Delta$dksA, but neither cross-
complementation could rescue the defect in $P_{\text{virB}}$ induction
(Fig. 1). Introduction of the SpoT overexpression plasmid
in a wild-type strain did not affect $P_{\text{virB}}$ activity compared with wild-type after 48 h
(Fig. 1, $P$ value $\leq 0.01$). Interestingly, time-course analysis
of DksA or SpoT overexpression in wild-type, $\Delta$spoT or
$\Delta$dksA bacteria did not reveal any change in the kinetics of
$P_{\text{virB}}$ induction (Fig. S2). Together, these results indicate
that the contribution of both spoT and dksA is necessary for
$P_{\text{virB}}$ induction, but neither factor is sufficient to activate this
promoter.

The SpoT-dependent $P_{\text{virB}}$ induction correlates with
ppGpp accumulation

Bartonella henselae SpoT belongs to the bifunctional
(p)ppGpp synthetase/hydrolase Rel [HS] family of proteins
(Atkinson et al., 2011) and sequence analysis indicated
that the residues which are essential for both enzymatic
activities (Hogg et al., 2004) are conserved in B. henselae
(data not shown). To confirm the involvement of B. hense-
lae SpoT in (p)ppGpp synthesis, we determined the cellular
concentration of ppGpp in wild-type, $\Delta$spoT and $\Delta$spoT/
spoT strains after 48 h of pH-dependent $P_{\text{virB}}$ induction in
M199/10% FCS. In contrast to wild-type B. henselae, no
ppGpp could be detected in a $\Delta$spoT mutant, whereas
ectopic expression of the deleted gene restored production
of the second messenger (Fig. 2A). The correlation
between the absence of ppGpp and an impaired $P_{\text{virB}}$
induction in the $\Delta$spoT mutant (Fig. 2B) indicated that the
accumulation of ppGpp is required for $P_{\text{virB}}$ induction and
that SpoT is responsible for the production of this second
messenger. The requirement of SpoT enzymatic activity
was further supported by the fact that expression of a
synthase mutant of SpoT [D259N substitution (Hogg et al.,
2004; Boutte and Crosson, 2011)] failed to restore the $P_{\text{virB}}$
inducibility in a $\Delta$spoT strain (data not shown). Noteworthy,
the downregulation of $P_{\text{virB}}$ expression observed in $\Delta$batR
and $\Delta$dksA mutants did not correlate with significant changes
in ppGpp levels (Fig. 2A and B), indicating that these
factors do not influence SpoT activity.

The effect of SpoT and DksA is dependent on the
response regulator BatR

To confirm the involvement of SpoT and DksA in the
regulation of the B. henselae VirB/D4 T4SS we performed

immunoblot analysis of crude bacterial extracts using an
antiserum directed against VirB5 (Quebatte et al., 2010), a
protein component of the VirB/D4 T4SS. Analysis of VirB5
expression in bacteria harvested after 48 h of host cell-free
induction confirmed that spoT (Fig. 3A) and dksA (Fig. 3B)
are required for the induction of the VirB/D4 T4SS, since
each mutant displayed a reduction in VirB/D4 levels similar
to the one observed in a $\Delta$batR mutant. To address the
hierarchical relationship between the SR components
SpoT and DksA and the response regulator BatR we
generated $\Delta$batR$\Delta$spoT and $\Delta$batR$\Delta$dksA double
mutants and tested them similarly for their ability to induce the
expression of the VirB/D4 T4SS. Indeed, both double
mutants were also impaired in VirB5 production (Fig. 3A and
B). Moreover, both SpoT overexpression in a $\Delta$batR$\Delta$spoT
mutant and DksA overexpression in a $\Delta$batR$\Delta$dksA strain failed to restore $P_{\text{virB}}$ induction, demonstrat-
ing that BatR is required for the action of DksA and
SpoT on the VirB/D4 T4SS expression.

SpoT and DksA are required for the expression of the
response regulator BatR

We have previously shown that the batR gene is induced
during the course of endothelial cell infection by B. hense-
lae (Quebatte et al., 2010). To better understand the
epistasis between the SR components SpoT and DksA and
the BatR/BatS TCS we addressed the question whether a
deletion in spoT or dksA would also influence BatR expres-
sion. To this end, we performed an immunoblot analysis of

© 2013 John Wiley & Sons Ltd, Molecular Microbiology, 90, 756–775
BatR expression in a set of mutants and complemented strains using an anti-serum directed against the BatR protein (Québatte et al., 2010). This analysis revealed that deletion of either spoT or dksA resulted in a strong decrease in BatR levels, which was largely complemented by plasmid-based expression of the deleted gene (Fig. 3C). This indicated that SpoT and DksA are required for the expression of the response regulator BatR and thus raises the question whether their contribution to the VirB/D4 T4SS expression would be the sole result of their control of BatR levels we performed a rescue experiment by introducing a plasmid encoded copy of the control of BatR expression would be through an independent mechanism.

**BatR overexpression exhibits opposite effects on spoT and dksA mutants**

To investigate whether the only contribution of DksA and SpoT to the VirB/D4 T4SS expression would be through the control of BatR levels we performed a rescue experiment by introducing a plasmid encoded copy of batR under the control of the ptalac promoter (pbatR (Québatte et al., 2010)) in ΔspoT and ΔdksA mutants and analysed the effect of BatR overexpression on PvirB activity by flow cytometry. Interestingly, BatR overexpression elicited a different response in ΔspoT and ΔdksA. Whereas BatR overexpression in ΔdksA was sufficient to restore a significant level of PvirB induction ($P < 0.01$), the same approach failed to rescue the ΔspoT mutant (Fig. 4A). On the contrary, BatR overexpression in ΔspoT had a negative effect on the residual PvirB activity still detected in this background (Fig. 4B). The differential effect of BatR overexpression in ΔspoT and ΔdksA mutants was confirmed by immunoblot analysis (Fig. 4C). Although BatR overexpression in the ΔdksA mutant typically resulted in lower accumulation that in ΔspoT background, this level was sufficient to restore some induction of the VirB/D4 T4SS, shown here by the accumulation of VirB5 protein.

Taken together, these results indicate that the effect of DksA on the expression of the VirB/D4 T4SS can be largely explained by its requirement for BatR expression. In contrast, as the requirement of SpoT could not be compensated by BatR overexpression, we hypothesized that (p)pGpp also contributes to PvirB activity by a BatR-independent mechanism.

**Fig. 3.** SpoT and DksA are required for the expression of the VirB/D4 T4SS and the response regulator BatR. A and B. Immunoblot analysis of VirB5 expression in cellular extracts of B. henselae after 48 h induction in M199/10% FCS supplemented with 500 μM IPTG (+) or without (-) using a polyclonal anti-serum directed against VirB5. A. Comparison of B. henselae wild-type and ΔbatR to ΔspoT and ΔbatRΔspoT mutants with (+) or without (-) complementation with pspoT. B. Comparison of wild-type and ΔbatR to ΔdksA and ΔbatRΔdksA mutants with (+) or without (-) complementation with pdksA. C. Immunoblot analysis of BatR expression in cellular extracts of B. henselae after 48 h induction in M199/10% FCS supplemented with 500 μM IPTG (+) or without (-) using a polyclonal anti-serum directed against BatR. Comparison of B. henselae wild-type to ΔspoT and ΔdksA mutants with (+) or without (-) complementation with pspoT or pdksA.

**Fig. 4.** The VirB/D4 T4SS inducibility can be restored by BatR overexpression in ΔdksA but not in ΔspoT. A. Expression of the chromosomally encoded PvirB:gfp reporter after 48 h of induction in M199/10% FCS supplemented with 500 μM IPTG (black bars) or without (white bars). GFP expression levels were determined as described for Fig. 1 for B. henselae wild-type or ΔspoT and ΔdksA mutants with (+) or without (-) complementation with pbatR. Values represent the mean ± SD from three independent experiments normalized to wild-type expression level. P values were calculated with two-tailed Student’s t-test ($*P < 0.01$). B. Representative flow cytometry data for the samples presented in A in the presence of 500 μM IPTG, displayed as double logarithmic dot-blot with side light scatter (SSC-H) and GFP intensity (FL1-GFP) values. The dashed line is set for comparison purpose. C. Immunoblot analysis of BatR and VirB5 expression in cellular extracts of B. henselae after 48 h induction in M199/10% FCS supplemented with 500 μM IPTG (+) or without (-) in the presence (+) or absence (-) of the complementation plasmid pbatR.
Deletion of \( \Delta dksA \) but not \( \Delta spoT \) abolishes invasome formation upon endothelial cell infection

The overall discrepancies between \( \Delta spoT \) and \( \Delta dksA \) with respect to \( P_{\VirB} \) regulation prompted us to assess the ability of \( \Delta spoT \) and \( \Delta dksA \) mutants to trigger the formation of invasomes. This characteristic F-actin rearrangement observed in human endothelial cells upon \( B. henselae \) infection is dependent on the functional assembly of the VirB/D4 T4SS (Schmid et al., 2004; Lu et al., 2013). Infection with \( B. henselae \) wild-type resulted in the formation of invasomes in more than 90% of the cells whereas the \( \Delta batR \) mutant failed to trigger the formation of these structures (Fig. 5). This defect was complemented by ectopic expression of BatR in accordance with previous studies (Quebatte et al., 2010). Similarly, infection with a \( \Delta dksA \) strain only resulted in very few invasome positive cells, while complementation of \( \Delta dksA \) with \( pdksA \) – and to a lesser extent with \( pbatR \) – restored this ability (Fig. 5). This finding is in accordance with the measured \( P_{\VirB} \) activity (Figs 1 and 4). However, infection with a \( \Delta spoT \) mutant revealed that this strain is still able to trigger a marked level of invasome positive cells, indicating that the VirB/D4 T4SS is still expressed and active in at least a fraction of the mutant population. Whereas the mild defect in invasome formation observed for the \( \Delta spoT \) mutant was complemented by conjugation of \( psopT \), introduction of \( pbatR \) in the same background fully abolished the ability of this strain to trigger invasome formation (Fig. 5). This result correlates with the decrease in \( P_{\VirB} \) activity observed by flow cytometry upon overexpression of BatR in \( \Delta spoT \) (Fig. 4B). Together, these results show that expression and assembly of the VirB/D4 T4SS depends on DksA, a requirement that is suppressed by overexpression of the response regulator BatR. In contrast, and despite a drastic decrease measured for both promoter activity (Fig. 1) and protein levels (Fig. 3A), a \( \Delta spoT \) strain still exhibits a VirB/D4 T4SS activity sufficient to trigger invasome formation upon host cell infection. This activity is dependent on BatR as invasome formation is fully abolished in a \( \Delta batR \Delta spoT \) double mutant (Fig. 5). We further deduced a dose-dependent effect of BatR in the absence of \( spoT \), based on the observed attenuation in the invasome formation ability of a \( \Delta spoT \) mutant upon BatR overexpression.

DksA requirement at \( P_{\VirB} \) is dependent on (p)ppGpp

Considering the differential effect of \( spoT \) and \( dksA \) deletions on the expression and assembly of the VirB/D4 T4SS we investigated the epistasis between both factors by constructing a \( \Delta spoT \Delta dksA \) double mutant and assessed its ability to trigger invasomes, as this read-out appeared to be the most sensitive in distinguishing the single-deletion strains. To our surprise, the \( spoT \) deletion showed an epistatic phenotype over \( dksA \) as the \( \Delta spoT \Delta dksA \) mutant was able to trigger marked levels of invasomes upon endothelial cell infection, although to a lesser extent than a \( \Delta spoT \) strain (Fig. 5). Overexpression of SpoT or DksA in the double mutant did not affect much this ability.

---

© 2013 John Wiley & Sons Ltd, *Molecular Microbiology*, 90, 756–775

---

**Fig. 5.** Differential effect of \( spoT \) and \( dksA \) deletion on invasome formation upon infection of human endothelial cells. Invasome quantification in human endothelial cells (Eahy.926) after 48 h of infection with different strains of \( B. henselae \) at a multiplicity of infection (moi) of 200. Infections were performed in M199/10% FCS with 500 \( \mu \)M IPTG (black bars) or without (white bars). Values represent the mean ± SD from three independent experiments. Asterisks indicate statistically significant difference to wild-type infection without IPTG (\( P < 0.01 \)) as determined by one-way ANOVA and Tukey’s post hoc tests.
of the expression of the chromosomally encoded \( P_{\text{virB}}:gfp \) reporter in the \( \Delta \text{spoT} \Delta \text{dksA} \) mutant reflected the observed results (Fig. S3). Our epistasis analysis suggests that the DksA requirement for the expression and assembly of the VirB/D4 T4SS is dependent on the second messenger (p)pGpp and that in absence of the latter DksA is dispensable.

**Overexpression of the alternative sigma factor RpoH1 restores expression of the VirB/D4 T4SS in \( \Delta \text{spoT} \) and \( \Delta \text{dksA} \) mutants in a BatR-dependent manner**

The SR components DksA and SpoT can influence gene regulation through a broad range of mechanisms, one of which is the modulation of the interaction between alternative sigma factors and the core RNAP by increasing the pool of free RNAP, a phenomenon also known as sigma competition (Österberg et al., 2011). The genome of *B. henselae* (Alsmark et al., 2004) encodes only four bona fide sigma factors (Ulrich and Zhulin, 2010): the housekeeping sigma factor RpoD (BH11530, \( \sigma^{70} \) family) and three alternative factors: a member of the extracytoplasmic function (ECF) family of sigma factors (RpoE/BH13830) and two uncharacterized members of the \( \sigma^{32} \) family (RpoH1/BH15210 and RpoH2/BH03810). Here we need to specify that we adopted the RpoH nomenclature of *Brucella melitensis* (Delory et al., 2006), which encodes the closest homologues of *B. henselae* RpoH1 and RpoH2 so far characterized. To assess whether the defect in \( P_{\text{virB}} \) induction in a \( \Delta \text{dksA} \) mutant (and to a lesser extent in \( \Delta \text{spoT} \)) would be the result of an alteration in sigma competition, we cloned and overexpressed each of these four sigma factors in wild-type, \( \Delta \text{spoT} \) and \( \Delta \text{dksA} \) strains and monitored their effect on \( P_{\text{virB}} \) activity by flow cytometry (Fig. 6A and B). In \( \Delta \text{spoT} \) and \( \Delta \text{dksA} \) backgrounds, the most striking effect was achieved by overexpression of the alternative sigma factor RpoH1 (BH15210), which markedly restored the induction of the *virB* promoter (Fig. 6A and B). The reactivation of \( P_{\text{virB}} \) induction by RpoH1 in \( \Delta \text{spoT} \) and \( \Delta \text{dksA} \) mutants correlated with accumulation of VirB5 proteins as demonstrated by immunoblot analysis (Fig. 6C), although the rescue of the \( \Delta \text{dksA} \) phenotype was more effective than the one of \( \Delta \text{spoT} \) under the tested conditions. RpoD overexpression also resulted in an increase of \( P_{\text{virB}} \) activity in \( \Delta \text{dksA} \) but not in \( \Delta \text{spoT} \). This effect was, however, not confirmed at the protein level (data not shown). In contrast, neither RpoH2 (BH03810) nor RpoE (BH13830) overexpression influ-
enced $P_{\text{virB}}$ activity in $\Delta\text{spoT}$ and $\Delta\text{dksA}$ mutants (Fig. 6A and B). In the wild-type background, overexpression of RpoH1 stimulated the activity of $P_{\text{virB}}$ about 1.5-fold, whereas overexpression of RpoD resulted in a bit less than a twofold decrease in $P_{\text{virB}}$ activity after 48 h of induction, suggesting an antagonistic effect between both sigma factors on the regulation of this promoter (Fig. 6A and B). In contrast, overexpression of RpoH2 and RpoE did not result in any observable phenotype in the wild-type background (Fig. 6A and B).

We then assessed whether the contribution of RpoH1 to $P_{\text{virB}}$ induction is BatR dependent, considering that BatR was required for the effect of both SpoT and DksA (Fig. 3A and B) and that BatR levels were affected in $\Delta\text{spoT}$ and $\Delta\text{dksA}$ mutants (Fig. 3C). Overexpression of RpoH1 in $\Delta\text{batR}\Delta\text{spoT}$ and $\Delta\text{batR}\Delta\text{dksA}$ double mutants failed to restore $P_{\text{virB}}$ induction, as did RpoH1 overexpression in the $\Delta\text{batR}$ mutant (Fig. S4). Of note, the $P_{\text{virB}}$ activity measured in a $\Delta\text{batR}\Delta\text{spoT}$ or $\Delta\text{batR}\Delta\text{dksA}$ strains was lower than the activity measured in a $\Delta\text{batR}$ single-deletion mutant ($P < 0.01$), indicating a possible contribution of SpoT and DksA to the BatR-independent activity of $P_{\text{virB}}$ (Fig. S4). Together, these results demonstrate that overexpression of the alternative sigma factor RpoH1 can restore the induction of the VirB/D4 T4SS in $\Delta\text{spoT}$ and $\Delta\text{dksA}$ strains and that the action of RpoH1 on $P_{\text{virB}}$ is dependent on BatR.

**Results**

**Deletion of rpoH1 abrogates the production of B. henselae VirB/D4 T4SS and prevents invasome formation upon endothelial cell infection**

To confirm the role of the alternative sigma factor RpoH1 (BH15210) in the control of *B. henselae* VirB/D4 T4SS expression we generated an *rpoH1* deletion mutant and tested the effect of this deletion on $P_{\text{virB}}$ activity by flow cytometry. In concordance with the overexpression experiments (Fig. 6A), *rpoH1* deletion resulted in a decrease of $P_{\text{virB}}$ activity (Fig. 7A). Interestingly, the expression profile of the $\Delta\text{rpoH1}$ mutant was reminiscent of class II mutants (e.g. $\Delta\text{batR}$), with a single-peak distribution of GFP fluorescence values reaching an intermediate intensity between the on- and the off-state observed in wild-type (Fig. 7A and Fig. S5B–C). RpoH1 overexpression complemented the loss of $P_{\text{virB}}$ induction in the $\Delta\text{rpoH1}$ mutant whereas BatR overexpression in the same background did not. On the contrary, introduction of the BatR overexpression plasmid in the $\Delta\text{rpoH1}$ mutant decreased the RpoH1-independent $P_{\text{virB}}$ activity measured in this background, similarly to the effect of RpoH1 overexpression in a $\Delta\text{batR}$ mutant (Fig. 7A). This prompted us to generate a $\Delta\text{batR}\Delta\text{rpoH1}$ double mutant. Interestingly, the $P_{\text{virB}}$ activity measured in the resulting strain was lower than the one measured in either single-deletion mutant, supporting a contribution from either regulator to the residual $P_{\text{virB}}$ activity measured in the $\Delta\text{batR}$ or $\Delta\text{rpoH1}$ mutants (Fig. S5). We also attempted to delete the two other alternative sigma factors encoded in *B. henselae* genome. The second $\sigma^{32}$ family factor *rpoH2* (BH03810) appeared to be essential for *B. henselae* under the tested conditions, as the second recombination event (loop-out) resulted in reversion to the

---

*Fig. 7.* Deletion of *rpoH1* (BH15210) affects $P_{\text{virB}}$ inducibility. 
A. Expression of the chromosomally encoded $P_{\text{virB}}\text{gfp}$ reporter after 48 h of induction in M199/10% FCS supplemented with 500 μM IPTG (black bars) or without (white bars). GFP expression was determined as described for Fig. 1 for *B. henselae* wild-type (WT) or $\Delta\text{batR}$ and $\Delta\text{rpoH1}$ mutants with (+) or without (−) complementation with *rpoH1*. Values represent the mean ± SD from three independent experiments normalized to wild-type GFP expression level. Asterisks indicate statistically significant difference ($P < 0.05$) as determined by one-way ANOVA and Tukey’s post hoc tests. 
B. Representative flow cytometry data for the samples presented in (A) in the presence of 500 μM IPTG, displayed as double-logarithmic dot-blots with side light scatter (SSC-H) and GFP intensity (FL1-GFP) values. The dashed line is set for comparison purpose.
wild-type locus for all tested clones (data not shown). In contrast, the deletion of the ECF sigma factor rpoE (BH13830) was obtained, but flow cytometry measurements did not indicate any contribution to the regulation of the virB promoter (Fig. S5A).

We then assessed the requirement of RpoH1 for invasome formation upon infection of human endothelial cells as a read-out for functional assembly of the VirB/D4 T4SS. Concordantly with the measurement of $P_{\text{virB}}$ activity, the $\Delta rpoH1$ strain failed to trigger the characteristic actin-ring structure associated with invasome mediated bacterial uptake (Fig. 8). This defect was complemented to almost wild-type levels by RpoH1 overexpression. As predicted from $P_{\text{virB}}$ activity measurements, neither BatR overexpression in $\Delta rpoH1$ nor RpoH1 overexpression in $\Delta\text{batR}$ was sufficient to rescue the invasome formation ability of these mutants, confirming the dual dependence of the $B. henselae$ VirB/D4 T4SS on both RpoH1 and BatR. RpoH1 overexpression in a $\Delta\text{dksA}$ strain restored invasome formation to almost wild-type levels in agreement with our flow cytometry and immunoblot analysis (Fig. 6A and B) whereas to our surprise it resulted in a marked decrease of invasome positive cells when overexpressed in a $\Delta\text{spoT}$ or $\Delta\text{spoT}\Delta\text{dksA}$ mutant (Fig. 8).

Summarizing, these results demonstrate that, in addition to the response regulator BatR, the alternative sigma factor RpoH1 is necessary for the expression of a functional VirB/D4 T4SS and confirm that RpoH1 overexpression in a $\Delta\text{dksA}$ mutant can rescue the mutant phenotype. In contrast, the restored induction of the VirB/D4 T4SS expression by RpoH1 overexpression in a $\Delta\text{spoT}$ mutant (Fig. 6B) did not correlate with our read-out for VirB/D4 T4SS functionality. On the contrary, the reduction of invasome formation measured for $\Delta\text{spoT}$ and $\Delta\text{spoT}\Delta\text{dksA}$ mutant strains complemented with $\text{rpoH1}$ (Fig. 9) suggest a pleiotropic effect for RpoH1 in the absence of (p)ppGpp.

**Shotgun proteomics analysis reveals that only a specialized subset of proteins is under the dual control of BatR and RpoH1**

We were further interested in deciphering the relative contribution of SpoT, DksA, RpoH1 and the BatR/BatS TCS to the regulation of $B. henselae$ virulence factors and to understand how these regulators individually modulate the adaptive response of the bacteria to their host. To this end we assessed global changes in protein expression in total lysates of wild-type, $\Delta\text{batR}$, $\Delta\text{spoT}$, $\Delta\text{dksA}$, $\Delta\text{rpoH1}$ and $\Delta\text{batR}\Delta\text{rpoH1}$ bacteria after 48 h of host cell-free pH-dependent induction using a shotgun proteomic approach combined with label-free quantification of the detected proteins (Glatter et al., 2012). This analysis resulted in the detection of 1023 individual proteins, corresponding to 68.7% coverage of the 1488 proteins encoded by the $B. henselae$ genome (for more details see Experimental procedures). Protein expression was normalized to the corresponding levels in $B. henselae$ wild-type and the proteins showing at least a 1.5-fold up- or downregulation ($q$-value $\geq 0.05$) were considered to be differentially regulated. This resulted in a total of 591 individual proteins deregulated in either mutant. With more than 400 proteins matching our criteria for differential regulation, the deletions of spoT or dksA resulted in the largest alterations in protein expression profile compared with wild-type bacteria. In contrast, with respectively 55 and 72 proteins showing a differential regulation compared with wild-type, the effect of $\text{rpoH1}$ or $\text{batR}$ deletion was of lesser impact. Noteworthy, $\text{rpoH1}$ deletion mostly resulted in protein downregulation (47 out of 55 regulated proteins) whereas spoT, dksA and batR deletion elicited slightly more upregulation (244 out of 437, 227 out of 424 and 41 out of 72 regulated proteins respectively). Co-regulation analysis identified a restricted overlap of 14 proteins downregulated in each of the tested mutants (Fig. 9A and Table 1A). Strikingly, the composition of this overlap group was dominated by the components of the VirB/D4 T4SS machinery and its secreted effectors (Bep), represented by nine proteins together with the response regulator BatR and the chromosomally encoded GFP reporter under the control of $P_{\text{virB}}$. Besides a confirmation of the flow cytometry and immunoblot analyses, these results also highlighted the narrow range of proteins under the control of this specific combination of regulators. Our analysis further revealed that although RpoH1 levels are not affected in a $\Delta\text{batR}$...
mutant, its levels are strongly reduced in both ΔspoT and ΔdksA mutants (Table 1B), suggesting that its expression is under the control of the SR (stringent control). This fact is further supported by the finding that the RpoH1 regulon represents a subset of the SR regulon controlled by SpoT and DksA, as seen in the Venn diagram representation of the proteins differentially regulated in the individual mutants (Fig. 9A and B). Co-regulation in either ΔdksA or ΔspoT background was found for 46 of the 47 downregulated and 7 of the 8 of the upregulated proteins in a ΔrpoH1 strain. In contrast, although BatR levels are significantly reduced in all tested mutants, analysis of the BatR regulon reveals both overlapping and independent regulation compared with the RpoH1 and SpoT or DksA regulated proteins (Fig. 9 and Table S2). Hierarchical clustering of the 55 proteins differentially regulated in B. henselae wild-type and a ΔrpoH1 strain resulted in six clusters (Fig. 9C and Table S2), with clusters 1 and 3 comprising most of the proteins co-regulated in each of the tested mutants. Cluster 2 and 4 consist of 24 proteins downregulated in ΔspoT, ΔdksA and ΔrpoH1 but not in ΔbatR. In this mutant, these proteins were either not significantly regulated (10 proteins) or upregulated (14 proteins). These proteins, whose expression appear to be repressed at physiological pH represent interesting candidates for the adaptation of B. henselae to the non-physiological environment encoun-

Fig. 9. LC-MS analysis of B. henselae regulatory mutants using label-free quantification of the detected proteins. A and B. Venn diagram distribution of the proteins differentially downregulated (A) or upregulated (B) in each tested mutants compared with B. henselae wild-type. All strains were cultivated for 48 h in M199/10% FCS (n = 3). The criteria for differential regulation were ≥ 1.5-fold down- (A) or upregulation (B) with a q-value ≤ 0.05. The total number of regulated proteins is indicated in brackets next to each mutant’s name. C and D. Hierarchical clustering of the proteins differentially regulated in ΔrpoH1 (55 proteins, C) or ΔbatR (72 proteins, D) and corresponding expression in the other tested mutants. R: ΔbatR, T: ΔspoT, A: ΔdksA, H: ΔrpoH1 and RH: ΔbatRΔrpoH1 mutant. Proteins with assigned functions are indicated and the corresponding data set is presented in Table S2.
tered in the midgut of its arthropod host. Noteworthy, 21 of these 24 proteins were co-regulated between ΔrpoH1 and ΔrpoH1ΔbatR, indicating a dominant effect of RpoH1 on the expression of these proteins. We performed the same analysis on the 72 proteins differentially regulated in wild-type and ΔbatR strains, which resulted in seven protein clusters (Fig. 9D and Table S2), four of which largely overlapped with those described for the RpoH1 regulon: cluster 1/3 comprising proteins globally down-regulated in any tested mutant and clusters 2/4 proteins downregulated in any of the tested mutants and upregulated in a ΔbatR background. Cluster 7 consists of 10 proteins that show an upregulation in ΔbatR but not in ΔrpoH1. Interestingly, these proteins all belong to the prophage I of *B. henselae* (Berglund et al., 2009), which shows an even stronger induction in ΔspoT mutant compared with ΔbatR, suggesting an activation of this phage in the absence of (p)ppGpp [(p)ppGpp(0) conditions].

Together, these results confirm the hierarchical relationship between BatR, SpoT, DksA and RpoH1 as inferred from our epistasis analysis and clearly designate RpoH1 as a SR sigma factor in *B. henselae*. Moreover, our clustering analysis reveals the complex organization of the BatR and RpoH1 regulons, with a restricted set of co-regulated proteins and a larger number of factors for which both regulators show different effects.

### The VirB/D4 and the Trw T4SS of *B. henselae* are differentially regulated by SpoT and BatR

Our shotgun proteomics analysis revealed an unexpected SpoT dependence for the expression of the *B. henselae* Trw T4SS, with the five detected components (TrwD–G, TrwK) strongly upregulated in a ΔspoT mutant and to a lesser extent in a ΔdksA mutant. The expression of these proteins was not affected in a ΔrpoH1 mutant, but showed a slight downregulation in a batR mutant (Fig. 10A). The Trw T4SS is the second T4SS encoded in the *B. henselae* genome and in contrast to the VirB/D4 T4SS, it is required for the colonization of the intraerythrocytic niche (Seubert et al., 2003) but is dispensable for endothelial cell infection (Dehio, 2004). Expression of the Trw T4SS is controlled by KorA and KorB, a heterodimeric negative regulator system (Seubert et al., 2003). Strikingly, the detected level of KorB protein was strongly reduced in both ΔspoT and ΔdksA mutants compared with wild-type bacteria, whereas this level was higher in a ΔbatR mutant (Fig. 10A). This inverse correlation suggests a regulatory mechanism taking place upstream or at the level of the KorB repressor, whose levels control the expression of the Trw proteins. To validate the observed de-repression of the Trw T4SS expression, we introduced a GFP promoter probe carrying the promoter of *trwH* [p5-18 (Seubert et al.,...]}
2003) in *B. henselae* ΔbatR, ΔspoT and ΔdksA strains that did not carry the chromosomal P\_virB:GFP fusion (deletion in the background of *B. henselae* RSE247). We monitored the *trwH* promoter activity by flow cytometry after M199/10% FCS induction. In the wild-type parental strain (RSE247), an increase in GFP expression was observed between 6 h and 48 h of induction, whereas no accumulation was observed in a ΔbatR mutant. In the background of ΔspoT and ΔdksA mutants that did not carry the chromosomal P\_virB:GFP promoter. Values represent the mean ± SD from three independent experiments normalized to wild-type expression level after 48 h. Asterisks indicate statistically significant difference (P < 0.01) as determined by one-way ANOVA and Tukey’s post hoc tests using wild-type (48 h) as a reference.

### Discussion

For this study, we used the pH-dependent inducibility of the *virB* operon of *B. henselae* (Quebatte et al., 2010) as a read-out for the establishment of an infection-competent state that mimics the onset of host interaction. This approach allowed us to uncover a new activating pathway that together with the BatR/BatS TCS is necessary for the expression of a functional VirB/D4 T4SS and its secreted effectors during endothelial cell infection. We show that this pathway is mediated by the alternative sigma factor RpoH1 and is controlled by the stringent response (SR) components SpoT and DksA. A summary of the interactions between the components of this regulatory network is presented in Fig. 11.

**Convergent but different contribution of SpoT and DksA to the expression of the VirB/D4 T4SS**

Coupling the expression of virulence factors to the metabolic state constitutes a widespread component of the niche adaptation strategy for a bacterial pathogen (Brown et al., 2008; Eisenreich et al., 2010; McDonough and Rodriguez, 2012). The second messenger (p)ppGpp rep-
for the understanding of this differential action is the \( \omega \) subunit of the RNAP. This dispensable component of the RNAP is only necessary for (p)ppGpp-dependent transcription regulation \textit{in vitro} in the absence of DksA (Vrentas \textit{et al.}, 2005). Whether the \textit{B. henselae} homologue of GreA (BH11700, 49% identity to \textit{E. coli}) or the \( \omega \) subunit of the RNAP encoded directly upstream of \( \text{spoT} \) contribute to this transcription regulation process remains to be investigated.

The stringent alternative sigma factor RpoH1 is a critical host adaptation factor for \textit{B. henselae}

The RpoH family of alternative sigma factors (\( \sigma^{32} \) family) is typically involved in the control of the bacterial heat-shock response (Grossman \textit{et al.}, 1984; Arsène \textit{et al.}, 2000) and is usually found as a single copy per genome (Martínez-Salazar \textit{et al.}, 2009). In contrast to RpoS (Dong and Schellhorn, 2010) or RpoN (\( \sigma^{44} \) family) (Kazmierczak \textit{et al.}, 2005; Shingler, 2011), the \( \sigma^{32} \) family of sigma factors is not classically associated with regulation of virulence. The situation is somewhat different for the \( \alpha \)-proteobacteria, which lack any RpoS homologues (Roop \textit{et al.}, 2003). In these bacteria the \( \sigma^{32} \) family of sigma factors has been proposed to compensate for this absence (Dufour \textit{et al.}, 2012). Interestingly, in the genus \textit{Bartonella} as in many Rhizobiacea and several other \( \alpha \)-proteobacteria the \textit{rpoH} gene has been duplicated (Green and Donohue, 2006; Martínez-Salazar \textit{et al.}, 2009, and references therein). Although each copy typically retained the ability to complement — at least partially — the temperature sensitivity of an \textit{E. coli} \( \Delta rpoH \) mutant (e.g. Karlis \textit{et al.}, 1998; Delory \textit{et al.}, 2006; Green and Donohue, 2006), the duplication has been associated with a functional diversification in respect to input signal(s), transcriptional control and regulated genes (Delory \textit{et al.}, 2006; Green and Donohue, 2006; Martínez-Salazar \textit{et al.}, 2009; Barnett \textit{et al.}, 2012; Dufour \textit{et al.}, 2012).

This functional diversification is well illustrated by looking at the modulation of host association in the symbiotic bacteria \textit{Rhizobium etli} (Martínez-Salazar \textit{et al.}, 2009) and \textit{Sinorhizobium melliloti} (Oke \textit{et al.}, 2001; Ono \textit{et al.}, 2001). For these bacteria, one RpoH1 parologue is critical for successful host interaction whereas the second RpoH copy is fully dispensable (Table S5). Similarly, deletion of \textit{rpoH2} in the zoonotic pathogen \textit{B. melitensis} was associated with a much more drastic effect on virulence than the deletion of \textit{rpoH1} (Delory \textit{et al.}, 2006). Interestingly, in these three bacterial species the \textit{rpoH} copy associated with host interaction is also the one that retained the ancestral role in responding to elevated temperatures (Table S5). The data presented here indicate that the pathogen \textit{B. henselae} has evolved a different regulatory strategy for its host interaction than its close relatives.
among the Rhizobiales as it relies on the other RpoH parologue, RpoH1 (Table S5).

Although overexpression of RpoH2 did not show any effect on the expression of the VirB/D4 T4SS, we cannot rule out some contribution of this factor to the host adaptation process. The temperature sensitivity associated with an rpoH2 deletion in *B. melitensis* and other Rhizobiales (Delory *et al.*, 2006) may explain why we failed to delete this gene when growing *B. henselae* at 35°C. Further experiments using different growth temperatures may circumvent this apparent problem and allow us to investigate the possible role of RpoH2 in *B. henselae* host adaptation.

Diversification of the transcriptional control for species carrying two paralogues of *rpoH* has been described in *S. meliloti* (Sauviac *et al.*, 2007), *R. etli* (Martínez-Salazar *et al.*, 2009) or Rhodobacter sphaeroides (Nuss *et al.*, 2009; Dufour *et al.*, 2012). For these bacteria, the copy associated with the heat-shock response (*B. henselae* rpoH2 orthologue) is under the control of the housekeeping σ2 (RpoD), whereas the second copy (*B. henselae* rpoH1 orthologue) is controlled by RpoE, an ECF sigma factor. Interestingly, *B. henselae* RpoH1 levels appear to be dependent on the SR components SpoT and DksA whereas deletion of these factors showed only limited effect on RpoH2 levels. Moreover, RpoH1 was recently identified among the strongest upregulated proteins in response to BatR overexpression (Omasits *et al.*, 2013), indicating that these two factors cross-regulate each other.

A recent characterization of the only *Bartonella* ECF sigma factor in *Bartonella quintana* (Abromaitis and Koehler, 2013) showed that this factor (RpoE) is induced at 28°C and in response to high haemin concentrations, supporting a role in the adaptation to the arthropod host environment. Here we show that deletion of the *B. henselae* RpoE does not affect the RpoH1-dependent expression of the VirB/D4 T4SS. Rather, RpoE levels are strongly decreased in a ΔrpoH1 mutant, indicating a different hierarchical organization of sigma factors in the *Bartonella*, with a stringent control of RpoH1 levels by SpoT and DksA, which in turn may regulate RpoE expression.

Possible implications for *B. henselae* host adaptation strategy

In this study, we present a new activation pathway that is required together with the BatR/BatS TCS for the expression of the VirB/D4 T4SS and its secreted effector proteins (Fig. 11). The BatR/BatS mediated signal transduction, which was shown to be activated at physiological pH (pH 7.4), contributes to the discrimination between the mammalian host and the alkaline environment encountered in the midgut of the arthropod vector (Boudko *et al.*, 2001; Santos *et al.*, 2011). Yet, completion of the infection cycle within the mammalian host further requires distinction between different niches for which the pathogen encodes an alternative set of virulence proteins. For instance, the VirB/D4 T4SS and its secreted effectors are required for the colonization of the primary replicative niche, but are dispensable at the blood stage (Schulein and Dehio, 2002). This is in contrast to the Trw T4SS that is required for the establishment in the intraerythrocytic infection environment (Seubert *et al.*, 2003) but is dispensable for the colonization of the primary niche (Dehio, 2004). However, both T4SSs seem to require the activity of the BatR/BatS TCS for their expression. We thus propose that the distinction between the early and late infection stage in the mammalian host is modulated by the SR components SpoT and DksA, presumably through the sensing of relative nutrient availability encountered at the different stages of infection. We describe here a simplified three-step adaptation scenario for the colonization of the mammalian host. We propose that in the early stage of the mammalian host infection the SR is activated. This is possibly the response to a nutrient poor environment. Together with the physiological pH of the host, activation of the stringent sigma factor RpoH1 triggers the expression of the early virulence factors, such as the VirB/D4 T4SS, which are required for the colonization of the primary replicative niche (Schmid *et al.*, 2004; Eicher and Dehio, 2012). In a second step, bacteria acquire competence for the colonization of red blood cells, a stage during which vascular endothelial cells have been proposed to play an important role (Chomel *et al.*, 2009). In this compartment, *B. henselae* displays mixed adaptive features. While the bacteria show BatR- and RpoH1-dependent sustained expression of the VirB/D4 T4SS and its secreted effectors, induction of the Trw T4SS is also initiated, although only to a moderate level. This apparent contradiction could be explained by intermediate levels of (p)ppGpp, sufficiently high to induce RpoH1, yet low enough to allow the release of the KorA/KorB repression (Seubert *et al.*, 2003). Finally, upon seeding into the nutrient rich bloodstream, the Trw T4SS would be fully de-repressed allowing colonization of the intraerythrocytic niche, a state that shares similarities with the situation observed in vitro in a ΔspoT mutant ([p]ppGp situations). The weak and heterogeneous DksA-independent expression of the VirB/D4 T4SS system in the absence of (p)ppGpp would further allow the recurrent re-infection of the primary niche characteristic for the infectious cycle within the mammalian host (Chomel *et al.*, 2009). Accordingly, we can hypothesize that proteins which are upregulated in a ΔbatR mutant and downregulated in strains mutated in their SR components would represent good candidates for adaptation factors required at stages of the arthropod colonization where the SR is induced. In support of this assumption, the *B. henselae* haemin-binding protein C (HbpC) presents exactly these
regulation properties: it is induced in arthropod-like conditions and was proposed to confer resistance to the toxic haemin concentration found in the arthropod gut (Rodén et al., 2012).

The signal(s) relayed by the SpoT/DksA/RpoH1 have not yet been resolved for B. henselae. The SpoT homologues of the related bacteria S. melloti and Caulobacter crescentus have been shown to be non-responsive to amino acid starvation, but are sensitive rather to carbon or ammonium starvation (Boutte and Crosson, 2011; Krol and Becker, 2011). Moreover, B. henselae harbours a conserved TGS domain [ThrRS, GTPase and SpoT (Wolf et al., 1999)] in the C-terminal part of SpoT which mediates a fatty acid metabolism-dependent regulation via its interaction with the acyl-carrier protein in E. coli (Battesti and Bouveret, 2006). We can thus postulate that B. henselae SpoT activity is controlled by a combination of metabolic signals, whose identification constitutes an exciting focus for further investigation.

Experimental procedures

Bacterial strains and growth conditions

All bacterial strains and plasmids used in this study are listed in Table S3. The reporter strain MQBS28 harbouring a chromosomal insertion of the virB promoter (bp –366 to +21) in front of gfpmut [pvel:gfpgfp (Quebatte et al., 2010)] was constructed in the background of RSE247, a spontaneous streptomycin-resistant mutant of B. henselae ATCC 49882T. MQBS28 served as wild-type for this study, and all other mutants were generated in this background, unless otherwise indicated. B. henselae strains were grown at 35°C for 2 days on Columbia agar supplemented with 5% sheep blood (CBA) in a humidified atmosphere containing 5% CO2. E. coli strains were kept at 37°C on solid support [Luria–Bertani (LB) medium + 1.5% agar] or in liquid culture (LB, 200 r.p.m.). Plasmids were introduced into B. henselae by conjugation from E. coli strain J23150 using three-parental mating as described previously (Dehio and Meyer, 1997). Unless otherwise indicated, antibiotics and other supplements were used at the following concentrations: (i) B. henselae: 2.4 μg ml−1 chloramphenicol (Cm), 10 μg ml−1 gentamicin (Gm), 30 μg ml−1 kanamycin (Km), 100 μg ml−1 streptomycin (Sm) or 500 μM isopropyl β-D-thiogalactosidase (IPTG) and (ii) E. coli: 100 μg ml−1 ampicillin (Ap), 20 μg ml−1 Cm, 20 μg ml−1 Gm, 50 μg ml−1 Km or 1 mM 2,6-diaminopimellic acid (DAP).

Construction of strains and plasmids

DNA manipulations were performed according to standard techniques and all cloned inserts were DNA sequenced to confirm sequence integrity. Chromosomal insertions or deletions of B. henselae were generated by a two-step gene replacement procedure as described in Schulein and Dehio (2002). For complementation/overexpression experiments, selected genes were cloned into the plasmid pCD341 (Dehio et al., 1998) under the control of the taclac promoter with the Shine–Dalgarno sequence of pPG100 (Schulein et al., 2005). Before conjugation with pbatR (pDT024), all B. henselae strains were pre-conjugated with the empty plasmid pMMB206 (Morales et al., 1991) that constitutively expresses the LaqI. This step was required to circumvent the observed reduced conjugation frequencies, presumably resulting from BatR expression before establishment of the repression by LaqI (Quebatte et al., 2010). Detailed description for the construction of each plasmid is presented in Supporting information. The sequence of all oligonucleotide primers used in this study is listed in Table S4.

Himar1 transposon mutagenesis and screen for loss of Pvir activation

Transposon mutagenesis was performed using the suicide transposon vector pML001, carrying the Himar1 transposon (Tn), a Gm resistance marker and a hyperactive transposase. This vector was derived from pH506 (Saenz et al., 2007) and detailed construction is provided in Supporting information. Tn mutants were obtained by conjugation of pML001 into MQBS28. After 5–7 days, individual Tn mutants were picked, expanded on CBA Gm plates for 3 days and re-streaked to be grown for another 2 days. Host cell-free induction of the Pvel:gfpgfp reporter was performed in M199 (medium 199, Gibco, Invitrogen) supplemented with 10% FCS (fetal calf serum, Amimed) as follows. Bacteria were harvested from plates and diluted into M199/10% FCS to a final concentration of OD600 = 0.008, which corresponds to c. 8 × 106 cfu ml−1. Half millilitre of the bacterial dilution was incubated for 48 h at 35°C and 5% CO2 into 48-well plates (Corning). Bacteria were harvested by pipetting and expression of the Pvel:gfpgfp reporter was determined by flow cytometry (see below). Each mutant for which the phenotype was confirmed by an independent set-up was selected for further investigation.

Determination of promoter expression by flow cytometry

Bartonella henselae strains were grown on Columbia agar plates containing 5% defibrinated sheep blood (CBA) supplemented with appropriate antibiotics and grown in a humidified atmosphere at 35°C and 5% CO2 for 3 days followed by re-streaking on fresh CBA plates and growth for another 48 h. Bacteria were resuspended in M199/10% FCS, diluted to a final concentration of OD600 = 0.008 and incubated in 48-well plates in a humidified atmosphere at 35°C and 5% CO2 for 48 h unless otherwise indicated. Expression of the Pvel:gfpgfp promoter was measured by recording GFP fluorescence excited at 488 nm using a FACSCalibur flow cytometer (Becton Dickinson). Data analysis was performed using the FlowJo software (Tree Star). Expression levels of the Pvel:gfpgfp reporter are given as the geometric mean of the GFP fluorescence intensities normalized to wild-type levels.

Mapping of Himar1 insertion sites

The DNA sequence flanking the Himar1 transposon mutant insertion was determined using arbitrary PCR based on the method described in O’Toole and Kolter (1998). The first PCR
was performed using a combination of an arbitrary primer (ARB-1B or ARB-1C) and an oligonucleotide annealing on the right end of the transposon and oriented in the outwards direction (prML211). The 50 μl reaction contained 0.2 mM dNTPs, 0.2 μM of each oligonucleotide, 1 μl of B. henselae boiled colony and 2 μl of Taq DNA polymerase (NEB) in 1× ThermoPol Reaction Buffer. For the second amplification using the same conditions, 1 μl of a 1:100 dilution of the first PCR reaction was used as template in combination with a primer complementary to the 5′ sequence of the arbitrary primer (ARB1) and a nested primer annealing on the right end of the transposon (prML191). The amplification conditions where the following: 5 min at 94°C, 30 cycles of 15 s at 94°C, 30 s at 38°C and 90 s at 68°C followed by 5 min at 68°C for the 1st PCR and 5 min at 94°C, 30 cycles of 15 s at 94°C, 30 s at 56°C and 90 s at 68°C followed by 5 min at 68°C for the 2nd PCR. The reaction products of the second PCR were cleaned-up with the Wizard SV System (Promega) and analysed by DNA sequencing using prMQ1342. The sequences of the oligonucleotide primers used are listed in Table S4.

**Quantification of cellular concentrations of ppGpp by LC-MS/MS**

Each strain of B. henselae was inoculated in 2 × 20 ml M199/10% FCS at a final concentration of OD\textsubscript{600} = 0.075 in 150 cm\textsuperscript{2} tissue culture flasks and incubated for 48 h. Cultures were rapidly cooled on iced-water and harvested by centrifugation. Pellets were washed 1× with PBS and aliquots were taken for protein quantification and for determination of P\textsubscript{wet} activity by flow cytometry. Pellets were resuspended in 300 μl of extraction solution (acetonitrile/methanol/ddH\textsubscript{2}O, 2/2/1, v/v/v) and extraction was performed as described for cyclic-di-GMP (Spangler et al., 2010). Levels of ppGpp were measured by liquid chromatography-tandem mass spectrometry on a QTRAP 5500 mass spectrometer (AB SCIEX) coupled with a Series 200 HPLC System (Perkin Elmer Instruments). The ppGpp was detected in positive ionization mode via selected reaction monitoring (SRM). Liquid chromatography separation was achieved on a porous graphitic carbon (Hypercarb) column. A linear gradient from 96% solvent A (10 mM ammonium acetate, pH 10) and 4% solvent B (acetonitrile) to 60% solvent B over 8 min at a flow rate of 0.6 ml min\textsuperscript{-1} was applied. The column was re-equilibrated for further 4 min to reach the starting condition again. To determine the amount of protein in each sample, 800 μl of 0.1 M NaOH was added to the bacteria culture aliquot, and the protein concentrations were determined with the Protein Assay kit (Bio-Rad) after heating the sample at 95°C for 15 min. Measurements were repeated in triplicate and values were expressed as ppGpp per mg of protein.

**Immunoblot analysis**

Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) and immunoblotting for the detection of the proteins BatR and VirB5 was performed as described (Quebatte et al., 2010): B. henselae cells were harvested after 48 h of growth in M199/10% FCS, washed 1× in PBS, resuspended to an OD\textsubscript{600} of 16 and mixed with an equal volume of 2× Laemmli buffer. Of each sample, 10 μl were separated by 14% SDS-PAGE and transferred to a nitrocellulose membrane (Hybond-C Extra, GE Healthcare). Immunoblots were incubated with polyclonal rabbit sera raised against recombinant BatR (1:20 000) or VirB5 (1:50 000), followed by a 1:15 000 dilution of a goat anti-rabbit horseradish peroxidase-conjugated secondary antibody (GE Healthcare). Immunoblots were developed using LumiGLO chemiluminescent substrate (KPL) and imaged using an ImageQuant LAS 4000 (GE Healthcare).

**Infection assay and quantification of invasome formation**

Infection and staining procedures where adapted from Lu et al. (2013). In brief, Ea.hy926 cells were seeded at a density of 2500 cells per well in 96-well plates. Following overnight incubation, cells were washed once with 100 μl of M199/10% FCS and infected with B. henselae strains at a moi = 200 in 100 μl of M199/10% FCS and incubated at 35°C and 5% CO2 for 48 h. Following the incubation, cells were fixed with 3.7% paraformaldehyde/0.2 M HEPES solution (pH 7.4) and permeabilized with 0.1% Triton X-100 in PBS during 15 min. Staining of F-actin was performed using TRITC- or A547-labelled phalloidin (Sigma, c\textsubscript{final} = 0.25 μg ml\textsuperscript{-1} and Dyomics, c\textsubscript{final} = 1.5 U ml\textsuperscript{-1} respectively). Bacteria were stained using serum 2037 (polyclonal rabbit anti-B. henselae, 1:100) as primary antibody and a Cy5-conjugated goat anti-rabbit IgG secondary antibody (Dianova, Hamburg, Germany, 1:100). DNA was stained using DAPI (Roche, 0.1 mg ml\textsuperscript{-1}). Ninety-six-well plates were subjected to automated microscopy using ImageXpress Micro automated microscopes (Molecular Devices, Sunnyvale, CA, USA). In every well, nine sites were imaged in four different channels: DNA (DAPI), reporter P\textsubscript{wet}-gfp (GFP), actin (TRITC or A547) and total bacteria (Cy5). Invasomes were defined and counted by eye (≥ 300 cells per tested condition).

**Analysis of B. henselae proteins by LC-MS/MS**

**Whole-cell lysis.** A total of 10\textsuperscript{4} cells were lysed in 50 μl of lysis buffer (8 M urea, 0.1% RapiGest, 0.1 M ammonium bicarbonate) and disrupted by two cycles of sonication for 20 s (Hielscher Ultrasonicator). Protein concentration was determined by the BCA assay (Thermo Fisher Scientific) using a small sample aliquot.

**Lys-C/trypsin protein digestion.** Lysates were reduced with 5 mM TCEP at 37°C for 60 min, alkylated with 10 mM iodoacetamide in the dark for 30 min and quenched with 12.5 mM N-acetyl-cysteine. Proteins were digested by addition of Lys-C (Wako) at 37°C for 4 h (protein to Lys-C ratio 100:1), diluted to a final urea concentration of 1.5 M with 100 mM ammonium bicarbonate buffer and further digested by addition of trypsin (Promega) and incubation at 37°C for more than 15 h (protein to trypsin ratio 50:1). After digestion, the samples were supplemented with trifluoroacetic acid and HCl to a final concentration of 0.5% and 50 mM respectively. Peptides were desalted on C18 reversed phase spin columns according to the manufacturer’s instructions (Microspin, Harvard Apparatus), dried under vacuum and stored at −80°C until further processing.
LC-MS analysis and label-free quantification. LC-MS/MS analysis of digested and purified *B. henselae* lysates was performed on a dual pressure LTQ-Orbitrap Velos mass spectrometer connected to an electrospray ion source (both Thermo Fisher Scientific) with a few modifications of what was described recently (Glatter et al., 2012). In brief, peptide separation was carried out using an EASY nLC system (Thermo Fisher Scientific) equipped with a RP-HPLC column (75 μm × 37 cm) packed in-house with C18 resin (ReproSil-Pur C18-AQ, 3 μm; Dr Maisch GmbH, Ammerbuch-Entringen, Germany). A linear gradient from 95% solvent A (0.15% formic acid, 2% acetonitrile) and 5% solvent B (98% acetonitrile, 0.15% formic acid) to 28% solvent B over 90 min at a flow rate of 0.2 μl min⁻¹ was applied. The data acquisition mode was set to obtain one high-resolution MS scan in the FT part of the mass spectrometer at a resolution of 60 000 full width at half-maximum (at m/z 400) followed by MS/MS scans in the linear ion trap of the 20 most intense ions. The charged state screening modus was enabled to exclude unassigned and singly charged ions and the dynamic exclusion duration was set to 20 s. The ion accumulation time was set to 300 ms (MS) and 50 ms (MS/MS).

For label-free quantification, the generated raw files were imported into the Progenesis LC-MS software (Nonlinear Dynamics, Version 4.0) and analysed using the default parameters. MS/MS-data were exported directly from Progenesis LC-MS in mgf format. Forward and reverse sequences of the predicted proteome of *B. henselae* (Omasits et al., 2013) were searched against a decoy database using MASCOT. The search criteria were set as follows: full tryptic specificity was required (cleavage after lysine or arginine residues); three missed cleavages were allowed; carbamidomethylation (C) was set as fixed modification, oxidation (M) as variable modification. The mass tolerance was set to 10 ppm for precursor ions and to 0.6 Da for fragment ions. Results from the database search were imported into Progenesis LC-MS and the peptide false discovery rate (FDR) was set to 1% using the number of reverse hits in the data set. The final protein lists containing the sum of the peak areas of all identified peptides for each protein, were exported from Progenesis and further statistically analysed using an in-house developed R script (SafeQuant) (Glatter et al., 2012).

The mass spectrometry proteomics data were deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository (Vizcaino et al., 2013) with the data set identifier PXD000180 and DOI 10.6019/PXD000180.

Acknowledgements

We specially want to thank Alexander Harms and Dr Simone Eicher for helpful comments and critical reading of the manuscript. We also thank Dr Simone Eicher for assistance with image acquisition as well as Mario Emmenlander and Dr Pauli Rämö for assistance with image analysis. We are thankful to Annette Garbe for ppGpp measurements. We thank Marius Liesch for providing the unpublished plasmid pML001 and Dr Patrick Guye for plasmids pPG611 and pPG612. We acknowledge support by Grant 31003A-132979 from the Swiss National Science Foundation (SNSF) and Grant 51RT 0_126008 for the Research and Technology Development (RTD) project InfectX in the frame of SystemsX.ch, the Swiss Initiative for Systems Biology (both to C.D.).

References


© 2013 John Wiley & Sons Ltd, *Molecular Microbiology*, **90**, 756–775


Supporting information

Additional supporting information may be found in the online version of this article at the publisher’s web-site.
3.4. RESEARCH ARTICLE IV (published)

Conjugative DNA transfer into human cells by the VirB/VirD4 type IV secretion system of the bacterial pathogen *Bartonella henselae*

Gunnar Schröder¹, Ralf Schuelein¹, Maxime Quebatte and Christoph Dehio

¹These authors contributed equally


3.4.1. Statement of the own participation

My contribution to the research article IV started at the stage of the manuscript revision. I established a quantitative PCR approach to map of the integrated regions of the translocated plasmid DNA into the chromosome of the several eukaryotic cell lines isolated in this study and further determined the integration junctions for some of these cell lines using TAIL-PCR. Based on these results I generated the Fig. 2, Fig. S2 and table S2 of the manuscript. I also contributed to the revision of the main text and wrote the SI results section.
Conjugative DNA transfer into human cells by the VirB/VirD4 type IV secretion system of the bacterial pathogen Bartonella henselae

Gunnar Schröder1,2, Ralf Schuelein1,3, Maxime Quebatte, and Christoph Dehio4

Biozentrum, University of Basel, 4056 Basel, Switzerland

Edited by Patricia C. Zambronski, University of California, Berkeley, CA, and approved June 24, 2011 (received for review December 22, 2010)

Bacterial type IV secretion systems (T4SS) mediate interbacterial conjugative DNA transfer and transkingdom protein transfer into eukaryotic host cells in bacterial pathogenesis. The sole bacterium known to naturally transfer DNA into eukaryotic host cells via a T4SS is the plant pathogen Agrobacterium tumefaciens. Here we demonstrate T4SS-mediated DNA transfer from a human bacterial pathogen into human cells. We show that the zoonotic pathogen Bartonella henselae can transfer a cryptic plasmid occurring in the bartonellae into the human endothelial cell line EA.hy926 via its T4SS VirB/VirD4. DNA transfer into EA.hy926 cells was demonstrated by using a reporter derivative of this Bartonella-specific mobilizable plasmid generated by insertion of an eukaryotic egfp expression cassette. Fusion of the C-terminal secretion signal of the endogenous VirB/VirD4 protein substrate BepD with the plasmid-encoded DNA-transport protein Mob resulted in a 100-fold increased DNA transfer rate. Expression of the delivered egfp gene in EA.hy926 cells required cell division, suggesting that nuclear envelope breakdown may facilitate passive entry of the transferred ssDNA into the nucleus as prerequisite for complementary strand synthesis and transcription of the egfp gene. Addition of an eukaryotic neomycin phosphotransferase expression cassette to the reporter plasmid facilitated selection of stable transgenic EA.hy926 cell lines that display chromosomal integration of the transferred plasmid DNA. Our data suggest that T4SS-dependent DNA transfer into host cells may occur naturally during human infection with Bartonella and that these chronically infecting pathogens have potential for the engineering of in vivo gene-delivery vectors with applications in DNA vaccination and therapeutic gene therapy.

Results

Several pathogenic bacteria have adopted bacterial conjugation systems for interkingdom transfer of macromolecular substrates into eukaryotic target cells or the extracellular milieu. These pathogenesis-associated T4SSs transport protein/ssDNA complexes like in bacterial conjugation or are reduced to protein transfer systems (1). They even may function in bacterial pathogenesis independently of any noticeable substrate transfer (7, 8). The VirB/VirD4 system of the plant pathogen Agrobacterium tumefaciens is the only known pathogenesis-associated T4SS that naturally facilitates conjugation-like DNA transfer to eukaryotic host cells (9). This T4SS mediates the transfer of a complex of the VirD2 relaxase and a covalently attached single stranded transfer DNA (T-DNA) derived from the tumor-inducing plasmid into infected plant cells. With the aid of additional virulence proteins, i.e., the cotransferred ssDNA-binding protein VirE2, the T-DNA is protected and targeted to the nucleus, where it integrates into the plant genome, leading to the expression of T-DNA encoded genes, which ultimately results in tumor formation. In addition to numerous reports on structure/function analysis of this paradigmatic T4SS machinery and the development of an efficient gene delivery system for plant genetic engineering, the A. tumefaciens VirB/VirD4 T4SS was also reported to be capable of transforming human cells under nonphysiological laboratory conditions (10). However, this plant cell-targeting DNA-delivery system is adapted to function at ambient temperatures but is entirely nonfunctional at human body temperature and therefore hitherto has failed to find any application in DNA vaccination or gene therapy.

The majority of the pathogenesis-associated T4SSs translocate bacterial effector proteins into mammalian host cells that subvert cellular functions to promote the host-associated lifestyle of the pathogen (11). Prominent examples are the Cag T4SS of Helicobacter pylori, which transfers the CagA effector protein into gastric epithelial cells, thereby mediating various cellular changes related to gastric disease (12); and the Dot/Icm T4SS of Legionella pneumophila, which transfers numerous effector proteins into amoebae or mammalian macrophages that facilitate the establishment of an intracellular bacterial replication niche (13). Of note, it has been demonstrated that some of these pathogenesis-associated T4SSs maintained their ancestral bacterial conjugation | transformation | bacterial infection | relaxase | orfT

Author contributions: G.S., R.S., M.Q., and C.D. designed research; G.S., R.S., and M.Q. performed research; G.S., R.S., and M.Q. analyzed data; and G.S., M.Q., and C.D. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

1G.S. and R.S. contributed equally to this work.
2Present address: European Patent Office, 10958 Berlin, Germany.
3Present address: Department of Microbiology and Immunology, University of Melbourne, Parkville 3039, Australia.
4To whom correspondence should be addressed. E-mail: christoph.dehio@unibas.ch.
This article contains supporting information online at www.pnas.org/cgi/content/full/1019074108/DCSupplemental.
Results

**B. henselae Translocates Plasmid DNA into Human Cells via the VirB/VirD4 System.** To test whether the VirB/VirD4 system of *B. henselae* may—in addition to its role in intracellular protein delivery—be capable of mobilizing DNA into human cells, we took advantage of the Bartonella-specific cryptic plasmid pBGR1, which, as its only discernable genetic elements, encodes a replication protein (Rep) and respective origin of replication (*oriT*), as well as a relaxase (Mob) and respective *oriT* (20). Although the Mob relaxase of pBGR1 does not contain a recognizable BID-like domain, we wondered whether this naturally occurring mobilizable plasmid might transfer into human cells during the course of infection. To monitor plasmid transfer, we introduced into *B. henselae* WT the plasmid pRS117, which is a pBGR1 derivative carrying an *egfp* reporter cassette under control of the CMV promoter that would mediate eukaryote-specific expression of *egfp* (Fig. 1A). As negative controls, we included the T4SS-deficient ΔVirB4 strain carrying pRS117, as well as the WT strain carrying a derivative of pRS117 in which the *mob* gene is disrupted (pRS086). Noteworthily, despite the deficiency for effector protein secretion, the ΔVirB4 strain does not display reduced viability during infection of the human endothelial cell line EA.hy926 (21). After 3-d infections of EA.hy926 cells with the different *B. henselae* strains, the cell populations were analyzed by flow cytometry to determine the percentages of GFP-positive cells (gpc) as functional readout for gene transfer and expression of the eukaryote-specific *egfp* reporter cassette (Fig. 1B and C). DNA transfer was detectable in cells infected with the WT strain carrying pRS117 (0.02% gpc), whereas infection with both control strains did not give rise to any detectable transfer event (detection limit, 0.003%). These data clearly demonstrate the transfer of the reporter plasmid pRS117 by *B. henselae* into human cells, and that this DNA transfer process is dependent on a functional VirB/VirD4 T4SS as well as the Mob relaxase.

**Fusion of Type IV Secretion Signal from Native VirB/VirD4 Substrate to Mob Dramatically Increases Efficiency of DNA Transfer into Human Cells.** In an effort to increase transfer efficiency, we fused to the C terminus of Mob the previously characterized C-terminal secretion signal of the VirB/VirD4 effector protein BepD (19), which consists of a BID-domain and a positively charged C-terminal tail. Strikingly, infections with WT *B. henselae* carrying pRS122 encoding this Mob/BID fusion protein (Fig. 1A) showed an approximately 100-fold–increased transfer rate to EAhy.926 cells compared with the pRS117 reference plasmid encoding the native Mob relaxase (Fig. 1B–D). These results support a correlation between Mob transfer rates and gene transfer rates and confirm the role of the relaxase as a pilot protein for the translocation of the plasmid DNA via the VirB/VirD4 T4SS (Fig. 1E), similarly as already demonstrated for the VirD2 protein during the delivery of the VirD2/T-DNA complex into plant cells (22).

**Integration of Translocated Plasmid DNA into Host Cell Chromosome.** To test whether the DNA transferred by this VirB/VirD4-dependent mechanism can integrate into the EAhy.926 chromosome, we included a resistance marker (*neo*) to facilitate geneticin selection of stable transfected cell lines, resulting in plasmid pRS130. Southern blot analysis of the genomic DNA of eight different genetic-resistance cell lines (A–H) revealed that the entire pRS130 plasmid or parts of it had integrated into the genome (Fig. S1). Mapping of the transferred DNA with plasmid specific probes followed by PCR-based refinement revealed heterogeneity in the DNA integration patterns (Fig. 2, Table S1, and SI Results). In four cases, the integration was largely restricted to the plasmid region encoding the *neo* resistance marker (cell lines D, E, G, and H). In two established cell lines, evidence was found that a plasmid fragment spanning the *oriT* region had been integrated (cell lines B and F), which could reflect the transfer and integration of a plasmid concatemer, as results from rolling-cycle replication or the integration of the plasmid DNA subsequent to precise recircularization at the *oriT*. To investigate whether, as in the case of VirD2 and the bound 5′ end of the *A. tumefaciens* T-DNA, the Mob relaxase can protect the plasmid DNA proximal to the *oriT* (23, 24), we mapped some representative integration borders by using thermal asymmetric interlaced (TAIL)-PCR (10, 25). Sequence analysis of the 5′ integration border in cell line C revealed that the junction to chromosomal DNA has occurred next to the *oriT* of pRS130 (Fig. 2B). This junction is reminiscent of an illegitimate recombination event, as it contains a 47-bp filler DNA (24) that consists of an inverted repeat in the 22 bp upstream of the *oriT*, followed by 25 bp showing short stretches of homology to both pRS130 and the human chromosomal insertion site. The three other integration sites (cell lines A, D, and E) presented in Fig. S2 do not share these characteristics, as no sequence similarity to the *oriT* has been detected. In the four cell lines (A, C, D, and E) for which the junctions were determined, the integrations lie in unrelated sequences of chromosomes 22, 4, 2, and 7, respectively, and thus no preference for specific integration sites could be distinguished. Taken together, our results show that the Mob protein can protect the 5′ end of the transferred DNA and suggest a mechanism of integration similar to what has been reported for the VirD2-bound T-DNA of *A. tumefaciens* (23, 24).

**Expression of a Delivered egfp Gene Requires Cell Division.** Several processes may interfere with the expression of ssDNA delivered into the cytoplasm of EAhy.926 cells (Fig. 1E). In the case of the
A. *tumefaciens* T-DNA transfer system for plant cells, the cotransferred ssDNA-binding protein VirE2 protects the naked T-DNA from degradation and assists passage of membranes by its pore-forming activity (26). Additionally, VirE2 and the 5′-bound VirD2 carry nuclear localization signals (NLSs), which facilitate import through the nuclear pore complex (27–29). In case of *B. henselae* plasmid transfer to EAhy.926 cells, there is no evidence for the existence of such helping factors, and the Mob relaxase does not contain any noticeable NLSs. Thus, nuclear uptake of the ssDNA may exclusively occur upon disassembly of the nuclear envelope during mitotic cell division. As cell division of the Ea.hy926 cell line is contact-inhibited, we measured the percentage of gpc of cell monolayers infected at different stages of confluence. Our results indicate that the expression of the translocated DNA inversely correlated with the cell density and thus appears to depend on cell division (Fig. 3A). In contrast, protein transfer as measured by the previously reported Cre recombinase assay for translocation (19) positively correlates with cell densities and is thus independent of cell division (Fig. 3A). Time-course experiments support these notions: whereas protein transfer rates constantly increased over a period of 6 d of infection, the rates of DNA transfer/expression reached a steady state at day 3 or 4 when the monolayers reached confluence and cell divisions ceased (Fig. 3B). Together these results suggest that the transferred DNA enters the nucleus passively after disassembly of the nuclear envelope, although we cannot exclude the sole or additional requirement for active DNA replication—which occurs during cell division—to allow recombination/integration to happen.

**Discussion**

T4SSs are known to transport different macromolecular substrates into diverse target cell types and thus are considered to represent the most versatile class of bacterial secretion systems (11). However, naturally occurring transkingdom DNA transfer has been described only for the pathogenic *A. tumefaciens*–plant interaction. In this report, we describe T4SS-mediated DNA transfer from a human pathogenic bacterium into cultured human cells that possibly may occur naturally during human infection. We demonstrate here that a derivative of the *Bartonella*–specific cryptic plasmid pBGR1 (20) is transferred into human cells in dependency of a functional VirB/VirD4 T4SS and the plasmid encoded relaxase. This pathogenesis-associated T4SS has most likely evolved from an ancestral bacterial conjugation

![Image](image-url)
A. *tumefaciens* transformation process, as described for the VirD2 protein of can protect the 5′ end of the transferred DNA during the delivery system in the recipient host cells supports this assumption. In contrast to most other established gene delivery systems (32) previously (19), but also for gene therapy and vaccination. In this study, the lack of an efficient ssDNA protection and nuclear delivery system in the recipient host cells supports this assumption. Finally, our analysis demonstrated that the Mob relaxase can protect the 5′ end of the transferred DNA during the transformation process, as described for the VirD2 protein of *A. tumefaciens* (30, 31).

The adaptation of the *B. henselae* VirB/VirD4 T4SS to human cells and its versatility for DNA transfer increases its potential for an application not only for protein therapy as described previously (19), but also for gene therapy and vaccination. In contrast to most other established gene delivery systems (32–34), conjugative DNA transfer may allow the transfer of very large DNA segments such as bacterial artificial chromosomes. By drastically increasing DNA transfer efficiency via fusion of the C-terminal secretion signal of BepD to the Mob relaxase, we have made a step toward the development of a potential tool for in vivo gene therapy in humans. Improvements of this system by assisting the steps following gene delivery, for increasing the efficiency of nuclear targeting and genomic integration, would need to follow.

Materials and Methods

**Bacterial Strains, Cell Lines, and Growth Conditions.** *B. henselae* and Escherichia coli strains were grown as described (21, 35). Bacterial strains used in this study and their origins are listed in Table S2. The endothelial cell line Ea. hy926, cultured by fusion of human umbilical vein endothelial cells (43) and the lung carcinoma cell line A549 were cultured as described previously (36). The cell line Ea. hy926/pRS56-c#B1 was cultured as described previously (19).

**Plasmid Constructions.** DNA manipulations were carried out according to standard procedures (37). The plasmids used in this study are listed in Table 1. Plasmid DNA isolation and PCR purification were performed using Qiagen or Macherey-Nagel columns according to the manufacturer’s instructions. *E. coli* Novablu was used for cloning steps and the *E. coli* strain J2150 for plasmid mobilization to *B. henselae* (38). For construction of pRS117, the 1861 bp HindIII fragment of pWay19 (gift from the Molecular Motion Laboratory, Montana State University, Bozeman, MT) containing the egfp gene under the control of the CMV promoter was blunt-ended by using Klenow polymerase and ligated with HindIII cleaved pBGR1, a derivative of pBGR1 (20). The plasmid containing the egfp gene and the kan gene in the same orientation was selected, yielding pRS117. For construction of pRS122, a 567-bp fragment encoding the secretion signal of *B. henselae* BepD flanked by two Agel restriction sites, obtained by using pRS51 as template and the oligonucleotides pRS350 and pRS351 (Table S3), was ligated with pRS117, and the plasmid containing the fragment in the correct orientation resulting in a fusion with the mob gene was selected. For construction of pRS506, pCHF1 (20) was digested with SalI/XmnI/NcoI and the 2,541-bp SalI-XmnI fragment containing the disrupted mob gene was ligated with pRS122 (Sall/XmnI). pRS130 was obtained by introducing the 1514 bp SnaBI-EcoRV fragment of pRS56 (19) containing the neo gene into HindIII-cleaved pRS129. pRS129, in turn, was constructed by ligating a 2,990-bp Sall-
Results - Research article IV

Table 1. Plasmids used in this study

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Description</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>pBGRK-18</td>
<td>Derivative of pBGR1 endowed with the kan resistance gene and the ColE1 replicon of pk18</td>
<td>Present study</td>
</tr>
<tr>
<td>prRS117</td>
<td>Derivative of prRS117 wherein the secretion signal of bepD (aa 352–534) is fused to the 3’-end of mob</td>
<td>Present study</td>
</tr>
<tr>
<td>prRS506</td>
<td>Derivative of prRS122 wherein the mob gene is disrupted</td>
<td>Present study</td>
</tr>
<tr>
<td>prRS130</td>
<td>Derivative of prRS122 containing a egfp reporter gene and a neo gene conferring G418 resistance in transformed cell lines</td>
<td>Present study</td>
</tr>
<tr>
<td>prSS1</td>
<td>Cre vector encoding NLS::Cre::BepD (aa 352–534) reporter protein</td>
<td>19</td>
</tr>
</tbody>
</table>

PstI fragment of pCX-eGFP (39) containing an egfp expression cassette including the chicken β-actin promoter and the CMV-IE enhancer sequence, with prLRS122. The primer sequences are provided in Table S3.

Detection of DNA or Protein Transfer from B. henselae into Human Endothelial Cells. Infections of Ea.hy926 (to measure DNA transfer) and Ea.hy926/prRS56-c881 cells (to measure protein transfer) with B. henselae strains were performed as described (19, 21), and the multiplicity of infection (34) and duration of infection were as indicated in the text and figure legends. The highest convergent rates observed in the present work versus the rates obtained previously (19) are related to the use of M19/10% FCS growth medium instead of DMEM/10% FCS during the infection, resulting in a higher activation of the VirB/VirD4 T4SS (40). After infections, cells were trypsinized, recovered in growth medium, and 4 × 10^4 to 1 × 10^5 cells were analyzed by using a FACS Calibur flow cytometer (Becton Dickinson).

Establishment of Stably Transformed Cell Lines. After initial infection of Ea.hy926 cells with B. henselae carrying prRS130 (strain R5E851), the cell growth medium was supplemented with gentamicin (20 μg/mL) to kill the bacteria. Serial dilutions of the cells were then cultured in cell growth medium containing G418 (500 μg/mL), and clonal colonies were propagated in this medium for isolation of individual cell lines.

PCR Analysis of the Transgenic Cell Lines. Genomic DNA isolated from the eight cell lines analyzed by Southern blot hybridization (Fig. S1 and SI Materials and Methods) was used for PCR detection of 13 different probes scattered across the entire sequence of pRS130. Genomic DNA from untransformed Ea.hy926 cells and the same DNA spiked with purified pRS130 were used as negative and positive control, respectively. PCR was carried out in 96-well format in 25 μL volume per reaction containing 2.5 ng genomic DNA, 30 nM of each primer, and 1× Power SYBR Green PCR Mix (Applied Biosystems). Amplifications were carried out according to the manufacturer’s instructions by using a StepOnePlus instrument (Applied Biosystems) and were analyzed for presence/absence of the targeted sequence by using StepOne software (Applied Biosystems). Sequences of the primers used for the PCR analysis are provided in Table S3.


3.5. UNPUBLISHED RESULTS RELATED TO RESEARCH ARTICLE I

This section contains unpublished results that we feel cover important aspects related to the work presented in the research article I (section 3.1.2). These results were either excluded from the manuscript for space reason or were obtained after the publication of the manuscript as part of follow-up experiments. As these results do not constitute a coherent story on their own, the conclusions will be discussed at the end of each result sections. I produced all results presented in this section.

3.5.1. Genetic analysis of the batR-batS locus

In the research article I, our mutation analysis of loci encoding the BatR/BatS TCS was restricted to the in-frame deletion of the gene encoding the RR BatR, as numerous attempts to generate a deletion in the gene encoding the HK BatS remained unsuccessful. We therefore hypothesized that batS would be essential for B. henselae growth under the tested conditions. This explanation was not fully satisfactory since even the first of the two recombination steps required for the deletion of batS could not be obtained. However, this step should not affect BatS expression in the resulting strain. Bacterial TCSs are typically co-transcribed as a bicistronic mRNA [1,2]. Interestingly, whereas batR is up-regulated during endothelial cell infection, the expression level of batS remains constant throughout the infection (Fig. 3.1A), which suggest a different transcriptional regulation for the two genes. Moreover, considering the genetic organization of the BatR/BatS TCS with batS encoded 39 bp downstream from batR (Fig. 3.2A), it remained unclear whether batR deletion would also affect batS expression.

Deletion of batR has a polar effect on BatS

To assess a possible effect of batR deletion on the expression of the HK BatS, we analyzed BatS levels in B. henselae wild-type and ΔbatR strains grown for 48 h on Columbia blood agar (CBA) plates or after 48 h of endothelial cell infection. Immunoblot analysis using a polyclonal serum raised against the soluble cytoplasmic domain of BatS revealed that deletion of batR results in a reduction of BatS expression below detection levels (Fig. 3.1B). This indicated that batR deletion exhibits a polar effect on BatS protein levels.
Figure 3.1: Expression of the BatR/BatS TCS. (A) Expression levels of batR (black squares) and batS (open circles) in B. henselae wild-type during human endothelial cell infection. Expression was determined by qRT-PCR in total RNA preparation after 15 minutes and 6, 15, 24 and 48 h infection. Mean values plus SD are represented for three independent experiments normalized to the expression levels of rpsL and to wild-type plate-grown bacteria. (B) BatR and BatS expression levels determined by immunoblot analysis. Total extract of B. henselae wild-type or ΔbatR stains were analyzed after 48 h growth on CBA plate or 48 h of Ea.hy926 cell infection with polyclonal antibodies raised against the BatR and the BatS proteins.

Considering the organization of the genetic locus encoding BatR and BatS (Fig. 3.2A) and the differential expression of batR and batS during infection (Fig. 3.1A), we propose that the polar effect of batR deletion on BatS expression is the result of the deletion of a batS promoter within the batR coding sequence. These results suggest that BatS is not necessary for B. henselae growth under the tested conditions, as we did not observe any growth defect for the ΔbatR mutant [3]. Moreover, the complementation data of the ΔbatR mutant presented in the research article I imply that BatR overexpression can rescue the mutant phenotype even in the absence of the histidine kinase BatS, as this protein cannot be detected in a ΔbatR mutant. This result was further validated and constitutes the basis for the generation of the data presented in the research article II.

Generation of ΔbatS and ΔbatRΔbatS (ΔbatRS) deletion mutants of B. henselae

To further investigate whether batS would be essential for B. henselae growth, we created two new deletion plasmids - the first with homology regions designed to delete both batR and batS (pMQ009) and the second to delete batS (pMQ034) with a different 5’ homology region compared to the previous trials. Conjugation of both constructs in B. henselae wild-type and subsequent recombination allowed the isolation of the expected mutants, as confirmed by overspanning PCR (data not shown). These results thus
demonstrated that batS was not necessary for growth under the tested conditions. We tested the effect of these mutations on the regulation of $P_{\text{virB}}$ and $P_{\text{beqD}}$ - the two promoters for which direct regulation by BatR has been shown (research article I). To this end, the reporter constructs $pP_{\text{virB}}:\text{gfp}$ and $pP_{\text{beqD}}:\text{gfp}$ were introduced into the mutant strains and promoter activities were determined by flow-cytometry after host-free pH-dependent induction (M199/10%FCS, pH 7.4, 35°C and 5% CO₂ for 48 h) using B. henselae wild-type and batR as control. Our results clearly indicate that deletion of either batS or batR batS have the same effect as batR deletion on the regulation of $P_{\text{virB}}$ (Fig. 3.2B) and $P_{\text{beqD}}$ (Fig. 3.2C). These results thus indicated that both BatR and BatS are necessary for the induction of these promoters and that no other histidine kinase can compensate for the absence of BatS in a batS mutant under the tested conditions.

![Diagram](image)

**Figure 3.2: Mutation analysis of B. henselae batR-batS locus.** (A) Genetic organization of the batR-batS locus in B. henselae wild-type and in the different deletion mutants. (B and C) Representative histograms illustrating the effect of the different deletions in the batR-batS locus on $P_{\text{virB}}$ (B) and $P_{\text{beqD}}$ (C) activity after 48 h of host-free induction (pH 7.4).

We further assessed the different mutants in the batR-batS locus for VirB/D4 T4SS functionality. To this end we infected the human endothelial cell line Ea.hy926 and assessed the ability of the different mutants to trigger the characteristic F-actin rings
associated with invasome-mediated uptake of *B. henselae* [4,5]. In contrast to wild-type bacteria and alike a ΔbatR mutant, the ΔbatS and ΔbatRS strains failed to induce these actin rearrangements (Fig. 3.3) confirming the inability of these mutants to assemble a functional VirB/D4 T4SS and to translocate effector proteins into the host cells.

![Figure 3.3: Mutation analysis of the *B. henselae* batR-batS locus and its effect on invasome formation.](image)

Together, our analysis of the batR-batS locus indicates that neither BatR nor BatS are essential for *B. henselae* growth under the tested condition and that ΔbatR, ΔbatS and ΔbatRS mutants are equally affected in the regulation of the virB and the hepD promoters. Promoter analysis and infection experiments showed that these mutants are phenotypically indistinguishable in respect of their VirB/D4 T4SS activity under the tested conditions. Moreover, our results suggest a non-canonical organization of the batR-batS locus with the possible presence of a batS promoter within the batR coding sequence. The implication of this particular operon organization on the positive feedback regulation of the BatR/BatS TCS and on expression of the BatR regulon remains to be elucidated.

### 3.5.2. Initial characterization of batR transcriptional regulation

We have determined the BatR regulon by transcriptome analysis and further demonstrated direct binding of the response regulator BatR to the promoter regions of *virB*
and bepD (research article I). Yet, the regulation of BatR expression itself remained largely uncharacterized. Data from our transcriptome analysis indicated that batR is induced with a similar induction kinetic as the virB operon during endothelial cell infection, i.e. a rather slow but persistent induction. This induction could be the result of a positive feedback regulation, in which activated BatR would promote its own transcription, a common feature for TCSs [2].

**Initial characterization of P\textsubscript{batR} regulation**

As a first step to characterize the transcription regulation of BatR we generated a reporter probe consisting in the 5’ upstream intergenic region of batR fused to the gene encoding the red fluorescent protein dsRED [6]. We introduced this construct into *B. henselae* wild-type or batR strains carrying a chromosomal copy of the P\textsubscript{virB}:gfp fusion and measured the promoter activity of both P\textsubscript{virB} and P\textsubscript{batR} by using flow cytometry after M199/10%FCS induction (Fig. 3.4).

**Figure 3.4: BatR-dependent activation of P\textsubscript{batR}.** Histogram representation of the expression of the chromosomally encoded P\textsubscript{virB}:gfp (A) and the plasmid encoded P\textsubscript{batR}:dsRED (B) in *B. henselae* wild-type (blue) or batR mutant (red) determined by flow-cytometry after 48 h induction in M199/10%FCS. The results from three independent clones are displayed. The dashed lines indicate the median of both “ON” and “OFF” population in the wild-type. (C) Dot blot representation of GFP and dsRED levels indicating correlation between the expression of both reporters. The data correspond to the 1\textsuperscript{st} histogram presented in (A) and (B) for *B. henselae* wild-type and batR mutant.
The results revealed a bimodal distribution for $P_{batR}$ expression in *B. henselae* wild-type after 48 h of M199/10% FCS induction (Fig. 3.4B). This effect was more pronounced than the one observed for $P_{virB}$ expression in the corresponding samples (Fig. 3.4A). Noteworthy, the expression pattern originating from the $P_{batR}$ fusion consistently showed a higher variability in term of bimodal distribution for both parallel (Fig. 3.4B) or independent assays (data not shown). Similarly to what we observed for the $virB$ promoter (Fig. 3.4A), the expression of $P_{batR}$ was decreased in a *batR* mutant (Fig. 3.4B). Interestingly, deletion of *batR* did not result in a complete loss of $P_{batR}$ activity, but rather to an intermediate induction level comprised between the two populations observed in *B. henselae* wild-type. The dot-blot representation of both reporter levels (Fig. 3.4C) supports a positive correlation between the expression of the *batR* and the *virB* promoters for both BatR-dependent and BatR-independent transcription activities.

**Initial characterization of BatR binding at $P_{batR}$**

The results of our promoter probe fusion analysis supported a role for BatR in the activation of its own transcription. To assess whether BatR would bind to its own promoter, we designed a set of three overlapping radiolabeled probes covering the 443 bp upstream intergenic region of *batR* and one probe within *batR* coding sequence and tested these for BatR binding by electrophoretic mobility shift assays (EMSA, Fig. 3.5A and B). Our results showed a BatR-concentration dependent shift for each of the tested probes, demonstrating BatR binding to its own promoter. Surprisingly, incubation of BatR with the radiolabeled probe designed within *batR* coding sequence also resulted in a dose-dependent shift, suggesting the presence of a BatR binding site in this sequence (Fig. 3.5A, +1 to +147).
Together, this initial characterization of BatR transcriptional activation supports our model in which BatR regulates its own expression by a positive feed-back regulation mechanism. Indeed, we could show that BatR binds to its own promoter and that the P_{batR} transcription activity displays a characteristic BatR-dependency. The relevance of the finding that BatR shows affinity for its own coding sequence remains open. Our characterization of BatR binding to P_{virB} (see 3.4.1.4) or P_{bopD} (research article I) suggests a low degree of sequence conservation between different BatR binding motifs. Therefore we cannot exclude the presence of a BatR binding site within the BatR coding sequence. We have discussed the probable presence of a batS promoter within BatR coding sequence (3.4.1.1). However, indications from our RNA-seq analysis (research article II) would clearly place this promoter region downstream from the sequence selected for our EMSA experiment. We can nevertheless not exclude a role for this cryptic binding site, possibly related to the expression of BatS. Finally, the finding of a BatR-independent activity at the BatR promoter is perfectly in line with the situation described for canonical TCS, for which the expression is typically dependent on two promoters. A constitutive promoter of rather low strength, allowing the basal expression of the TCS which is necessary for initial signal recognition, and a second promoter that is positively controlled by the activated RR and which is required for the positive feedback regulation loop [1,2].

Figure 3.5: BatR binding to the batR promoter. (A, B) Four different radiolabeled DNA probes (2 to 4 fmol) spanning part of the upstream intergenic region or the coding region of batR were incubated with 0 (-), 0.8, 1.6, or 3.2 µg of purified BatR protein. The position of the probes relative to the ATG of the batR gene is indicated.
3.5.3. Differential induction kinetics between \( P_{\text{virB}} \) and \( P_{\text{bepD}} \) in response to BatR expression

Transcriptome profiling of \( B. henselae \) during human endothelial cell infection revealed a different kinetic of induction between the VirB/D4 T4SS and its secreted effector proteins, with the genes encoding the VirB/D4 machinery showing an earlier induction compared to the effectors (research article I). This property was further confirmed by quantitative PCR and by flow-cytometry using promoter-probe fusion. Having identified BatR as a direct activator of both \( P_{\text{virB}} \) and \( P_{\text{bepD}} \) we hypothesized that the different kinetics of induction could result from a different BatR binding affinity for both promoters. We took advantage of the finding that BatR overexpression results in the activation of its target promoters even in the absence of BatS (see section 3.4.1.1) and assessed our hypothesis as follows. In order to measure the transcriptional response of \( P_{\text{virB}} \) and \( P_{\text{bepD}} \) to BatR expression, we constructed two reporter strains carrying either the \( P_{\text{virB}}:\text{gfp} \) or the \( P_{\text{bepD}}:\text{gfp} \) fusion as a single copy insertion in the chromosome. We then monitored the promoter activity of these strains in response to BatR expression after introduction of a plasmid encoded copy of \textit{batR} under the control of an IPTG inducible promoter.

In the background of the wild-type and without ectopic expression of BatR both reporters displayed similar induction kinetics as previously determined by their plasmid encoded counter-parts. The \( \textit{virB} \) promoter was induced earlier and reached its maximal activity between 24 and 48 h, whereas \( P_{\text{bepD}} \) induction was delayed and still increased after 48 h of induction (Fig. 3.6). Introduction of the BatR overexpression plasmid in wild-type bacteria resulted in an attenuation of \( P_{\text{virB}} \) induction in the absence of IPTG (Fig. 3.6A, dotted line). This suggests that basal BatR expression from the \( P_{\text{lac}} \) acts negatively on the regulation of \( P_{\text{virB}} \). Overexpression of BatR in the wild-type background did not affect the induction kinetics of the promoter, although the promoter activity decreased after 48 h induction, unlike the strain without plasmid (Fig. 3.6A). In contrast to the \( \textit{virB} \) promoter, \( P_{\text{bepD}} \) was not affected by the BatR expression plasmid in the absence of IPTG (Fig. 3.6B, dotted line). However, it strongly reacted to BatR overexpression, reaching more than 60% GFP positive cells after only 6 h of IPTG induction, in contrast to less than 5% for the strain without plasmid. The percentage of GFP positive cells remained constant till 24 h after what it started to decrease.
To assess the potential cross-talk between the ectopically expressed BatR and the endogenous BatR/BatS TCS, we repeated the BatR overexpression experiment in the background of *B. henselae* ΔbatRS mutant. Globally, the results were similar compared to those obtained by overexpression of BatR in the wild-type strain. For the *virB* promoter, overexpression of BatR in the ΔbatRS background resulted in a very similar induction profile compared to wild-type bacteria without promoter (Fig. 3.6C) without affecting the kinetics of induction. This demonstrated complementation of the deletion phenotype by the sole overexpression of the response regulator BatR. In contrast, BatR overexpression triggered a strong activation of the *bepD* promoter already after 6h of induction, whereas no induction was observed in the absence of IPTG (Fig. 3.6D).

![Graphs showing GFP production](image_url)

**Figure 3.6:** Different response of P_{virB} and P_{bepD} to BatR overexpression. GFP production from the chromosomally encoded P_{virB} (A, C) and P_{bepD} (B, D) in *B. henselae* wild-type (A, B) or ΔbatRS mutant (C, D) determined by flow cytometry after 6, 12, 24 and 48 h in M199/10%FCS. Wild-type bacteria without plasmid (black square) are compared to bacteria carrying a BatR overexpression plasmid without addition of IPTG (dashed line, open square) or in the presence of 500 μM IPTG (continuous line, open square). Mean values plus SD are represented for a representative experiment set-up in triplicate.
Together, these results suggest a different architecture for the \textit{virB} and the \textit{bepD} promoter. The \textit{bepD} promoter displays a quick response to BatR overexpression, which is in line with a direct activation of P\textsubscript{bepD} by BatR. In contrast, although the \textit{virB} promoter can also be induced by BatR overexpression in the absence of the endogenous BatR/BatS TCS, its induction follows the same kinetic as the one observed in wild-type. Thus whereas a simple model for P\textsubscript{bepD} activation by BatR is enough to explain the induction kinetics of this promoter, our results indicate that one or several additional regulation mechanisms are involved in the control of P\textsubscript{virB}. This assumption was largely verified by the identification of the stringent alternative sigma factor RpoH1 as co-activator of P\textsubscript{virB} (\textit{research article III}). In the light of these results, the slow response of the \textit{virB} promoter to BatR overexpression may reflect the time required to mount the stringent response under the tested conditions.

\textbf{3.5.4. Refinement of BatR binding motif at P\textsubscript{virB}}

We have demonstrated direct binding of the response regulator BatR to the promoter regions of \textit{virB} and \textit{bepD} (see Fig. 6 and 7 in \textit{research article I}) and correlated this binding with the pH-inducibility of these promoters, demonstrating that BatR constitutes a transcriptional activator for the \textit{virB/D4} T4SS and its secreted substrates. Our analysis however failed to define the BatR binding motif on both regulated promoter. We further investigated BatR binding to P\textsubscript{virB} in order to better delineate the sequence requirement for BatR binding to this promoter.

\textit{Design of a minimal competitor for BatR binding to the P\textsubscript{virB} by EMSA}

The results presented in the \textit{research article I} (section 3.1.2) allowed us to estimate the location of the BatR binding site between bp -306 and bp -277 relative to the \textit{virB2} ATG. We aimed to refine this data by using shorter double stranded DNA fragments as competitor in our EMSA experiment. In a first approach, we designed 6 pairs of 22 bp overlapping double stranded oligonucleotides (1-6 in Fig. 3.7A and C) spanning the predicted BatR binding site and assessed their ability to compete BatR binding to a \textsuperscript{33}P-labeled P\textsubscript{virB} probe. None of these shorter competitors did interfere with BatR binding (Fig. 3.7A, lanes 1-6) in contrast to the previously described 50-mer competitor prDT087/088 (Fig. 3.7A, lane 7). We extended the size of the competitors to 30 bp and tested three new overlapping candidates in our EMSA set-up (Fig. 3.7B). Two of these
oligonucleotides competed with BatR binding to the $^{33}$P-labeled $P_{virB}$ probe (Fig. 3.7B), with the competitor “a” (prMQ1231/1232) showing strongest binding inhibition. These results allowed us to refine the location of the BatR binding site to a 30 bp region within $P_{virB}$. They also indicated that 22 bp are not sufficient to compete for BatR binding under the tested condition.

**Figure 3.7: EMSA competition assay.** (A, B) EMSA of the 203 bp fragment of the $virB$ promoter in the presence of unlabeled competitor. Two femtomoles of radiolabeled probe were incubated with either no (-) or 2 μg (+) of purified His$_6$-BatR protein in the absence (-) or presence of 1,000 X molar excess of double-stranded competitor. (A) EMSA in the presence of 22 bp competitors (1-6) or the 50 bp competitor prDT087/88 (7). (B) EMSA in the presence of 30 bp competitors (a, b, c) or the 50 bp competitor prDT087/88 (7). The relative localization of the competitors is depicted in (C) and their position relative to $virB2$ ATG is indicated on top of the panel. The sequence of the competitors is provided in Table 3.3

**Mutation analysis of the BatR binding site by EMSA competition assay**

To gain further insights into the residues important for BatR binding on the 30 bp of $P_{virB}$ that were delineated by competition experiments (Fig. 3.7) we generated 15 derivatives of competitor “a” (prMQ1231/1232), each carrying the sequential replacement of a dinucleotide by its complement counterpart (Table 3.3) We then tested these mutated competitors in our EMSA competition assay for their ability to prevent BatR binding to the radiolabeled $P_{virB}$ fragment. The results from a representative experiment are presented in Fig. 3.8. The residues of the putative BatR binding motif were classified into three
qualitatively categories, depending on the effect of their mutation in three independent replicates of the EMSA experiment: critical residues, whose mutation strongly reduced the competition ability of the double stranded oligonucleotides (mutations 5, 8, 11, 12, 13), neutral residues, whose mutation did not affect the competition (mutations 1, 2, 3, 7, 9, 10, 14) and residues, whose mutation resulted in an intermediate phenotype (mutations 4, 6, 7, 15). The results, displayed as color-coded nucleotides in Fig. 3.8C, summarize the sequence conservation required for BatR binding at P_{virB}. The highlighted sequence partially overlaps with the two indirect repeats present in the sequence (IR1 and IR2 in Fig. 3.8C).

Figure 3.8: EMSA competition assay with mutated competitors. (A, B) EMSA of the 203 bp fragment of the virB promoter in the presence of unlabeled competitor. Two femtomoles of radiolabeled probe were incubated with either no (-) or 2 μg (+) of purified His_{6} BatR protein in the absence (-) or presence of 1,000 X molar excess of un-mutated (a) or mutated (1-15) double-stranded competitor. (C) The sequence of the un-mutated 30 bp competitor is displayed in capital letters. The numbers on top of the sequence corresponds to the mutated dinucleotide (replaced by the complementary base) in the EMSA experiment displayed in (A) and (B). Two perfect indirect repeats (IR1) and (IR2) are indicated by arrows. Two additional nucleotides present directly downstream from the 30 bp motif are displayed in italics. The sequence of the competitors is provided in Table 3.3.
**Mutation analysis of the BatR binding site on P\textsubscript{virB} by promoter probe fusion**

To better characterize the involvement of the refined BatR binding motif for the regulation of the \textit{virB} operon, we analyzed the effect of the different dinucleotide mutations already tested in our EMSA competition assay on the reporter constructs pP\textsubscript{virB}:\textit{gfp}. We created 15 derivatives of pP\textsubscript{virB}:\textit{gfp}, each containing one of the mutations and introduced the resulting reporter constructs in either \textit{B. henselae} wild-type or \textit{batR} mutant. We then compared the promoter activity of each of these mutated derivatives by flow cytometry after incubation in M199/10\%FCS (Fig. 3.9A and B). As already described [3], expression of the un-mutated pP\textsubscript{virB}:\textit{gfp} in \textit{B. henselae} wild-type is characterized by a bimodal distribution, with a population of high expressers (“ON-state”) and a population of low expressers (“OFF-state”). The median level of both populations is displayed by the dotted lines (I) and (II) in Fig. 3.9. When introduced in a Δ\textit{batR} mutant, expression of the un-mutated pP\textsubscript{virB}:\textit{gfp} results in a unimodal distribution, with an expression level lower than the wild-type ON-state but higher that the OFF-state. The median level of this BatR-independent promoter activity is displayed by a dotted line (r) in Fig. 3.9.

In the background of \textit{B. henselae} wild-type the different mutations either showed no effect (mutations 1-4, 8-10 and 14-15) or resulted in the attenuation of the GFP expression levels (mutations 5-7 and 11-13) to a similar extent for each of these mutations. These results globally correlated with the results of the EMSA competition presented in Fig. 3.8, indicating that the decrease in transcription activity observed for the mutations 5-7 and 11-13 could result from an impaired binding of BatR to the mutated binding motif. Noteworthy, none of these mutations abolished the bimodality of GFP expression. Moreover, the promoter activity of any of these mutated P\textsubscript{virB} was clearly distinct from the promoter activity of the un-mutated plasmid expressed from a Δ\textit{batR} mutant (Fig. 3.9B, first row).
Results - Unpublished results related to research article I

Figure 3.9: Mutation analysis of the BatR binding motif in \( p_{virB}:gfp \). (A, B) Representative histograms of GFP expression from un-mutated (0) or mutated (1-15) \( p_{virB}:gfp \) in *B. henselae* wild-type (A) or ΔbatR mutant (B) after 48 h of M199/10% FCS induction at pH 7.4. The numbers displayed on the vertical axis (0-15) refer to the mutations in the BatR binding motif (Table 3.3). The un-mutated BatR binding site is depicted in the middle of the panel. The vertical dotted lines represent the median value for the OFF (I) and the ON (II) populations of the un-mutated promoter in *B. henselae* wild-type and the median value of the un-mutated promoter in the batR mutant (r). The histograms are color coded as follows: black: un-mutated plasmid, blue: neutral mutation, red: down mutation, green: gain of function mutation. (C) Color coded summary of the mutation analysis of the 30 bp BatR binding region from the EMSA competition assay (line 1) or form the analysis of \( p_{virB}:gfp \) expression in *B. henselae* wild-type (line 2) or batR mutant (line 3). (D) Conservation analysis of the BatR binding motif in the *virB* promoter of sequenced *Bartonella* spp. The conservation between the different aligned sequences is represented by a logo on top of the alignment. Only one BatR-binding site is displayed for all *Bartonella* species with more than one isolate sequenced, as they all showed 100% conservation (data not shown). (1) The BatR binding site was identical for the three copies of the P\(_{virB}\) present in these species so only one sequence was included in the alignment. (2) The BatR binding site was identical in two out of three copies of the P\(_{virB}\) present in these species. Only the two divergent sequences are indicated.
Unexpectedly, our mutation analysis also resulted in altered GFP expression levels when performed in the background of the batR mutant (Fig. 3.9B) indicating that some of the residues involved in the BatR-dependent activation of P_{virB} are also involved in its BatR-independent activity. All mutations showing an effect in B. henselae wild-type also decreased the promoter activity in the ΔbatR mutant, with one (mutation 7 in Fig. 3.9B) having the most drastic effect. In contrast, three mutations that did not show any effect in the wild-type background (mutations 4, 10 and 15) resulted in an enhance promoter activity in batR, promoting an expression level similar to the ON-state of the un-mutated promoter in the wild-type.

Finally, we looked at the conservation of this refined BatR binding site in the virB promoter of the different sequenced strains of Bartonella spp. Indeed, there was a trend towards a higher conservation for the residues that were important for the BatR-dependent activation of P_{virB} as determined by our functional assays for B. henselae (Fig. 3.9C). Preliminary characterization of a gfp fusion to the different virB promoters of B. claridgeiae confirmed their BatR-dependent expression in B. henselae (data not shown), suggesting a conserved regulation mechanism. However, whether all the observed variations would be permissive in B. henselae remains to be determined.

Concluding remarks

Summarizing, a combination of in vitro and in vivo assays allowed us to delineate the BatR-binding on P_{virB} promoter as a set of 2 x 6 nucleotides separated by a 6 bp spacer, a motif that is highly conserved in the virB promoter of different Bartonella spp. Members of the OmpR/PhoB subfamily of RRs [1], to which BatR belongs [3], display very different sequence requirements for DNA binding [7]. For example the DNA-binding motif of the RR PhoB is characterized by a 18-20 base pair (bp) pho box largely conserved among proteobacteria, consisting of two direct repeats of 7-11 bp separated by a 4 bp spacer region [8]. In contrast, the binding sites of the RR OmpR share only very limited homology [9] and OmpR is prone to unspecific DNA recognition. This was explained by the few specific contacts made between the protein and the DNA as observed by NMR spectroscopy [10]. We noticed the presence of a perfect inverted repeat partially overlapping the BatR binding site. This repeat is however unlikely to be part of the BatR binding motif as RRs of the OmpR/EnvZ subfamily recognize direct repeats on the DNA, as a result of the head-to-tail dimerization of their DNA binding domain in their activated form [11,12].
Results - Unpublished results related to research article I

As we failed to identify our refined BatR-binding motif in the promoter of any other BatR-regulated genes, including the bepD promoter (data not shown), we hypothesized that BatR binding to DNA may not require a high sequence conservation, similarly to what was described for OmpR [13]. This hypothesis is indirectly supported by the absence of DNA-binding motif description for the otherwise well characterized close homologues BvrR of Brucella spp. [14,15] and ChvI of Agrobacterium tumefaciens [16,17].

Interestingly, mutation analysis of the P\textsubscript{virB}:\textsuperscript{gfp} reporter in the background of a batR mutant revealed a partial overlap between the BatR-binding site and the residues whose mutation affected the BatR-independent transcription activity of P\textsubscript{virB}. These results suggest binding of the virB promoter by another – yet elusive – transcription regulator with a binding site overlapping the one of BatR. In B. henselae wild-type, the binding of this second factor would be occulted by BatR, explaining why the specific BatR-independent transcription profile was never observed in that strain. The data obtained by RNA-seq analysis (research article II) suggests a location of the transcriptional start site (TSS) of virB\textsubscript{2} 124 bp downstream from the batR box. This configuration, that remains to be validated by independent methods, speaks against an overlap between the BatR binding site and the -35 /-10 region of P\textsubscript{virB}, which could have constituted a explanation for the observed perturbation in the BatR-independent transcription activity, presumably related to the binding of the alternative sigma factor RpoH1 (see research article III). The mutations that resulted in a BatR-independent locked-on promoter (mutations 4, 10 and 15 in Fig. 3.7) constitute an interesting basis for follow-up experiments, if their effect can be validated in their native chromosomal context.
3.6. UNPUBLISHED RESULTS RELATED TO RESEARCH ARTICLE III

This section contains unpublished results related to the work presented in the research article III (section 3.3.2) and covers the initial characterization of some of the transposon mutants isolated in our screen aiming at identifying B. henselae factors required for the pH-dependent induction of the virB promoter. These results were not included in the research article III as they did not match the focus of the publication. I produced all results presented in this section with the exception of the BatR overexpression experiments (Fig. 3.13A) performed by Mathias Dick in the frame of his master thesis. The acquisition and the primary analysis of the mass spectrometry data were performed by Alexander Schmidt from the Proteomics Core Facility of the Biozentrum.

3.6.1. Functional characterization of B. henselae transposon mutants affected in P\textsubscript{virB} regulation

Our screen aiming at identifying genes involved in the pH-dependent induction of the virB promoter resulted in the isolation of 53 mutants mapping to 39 individual genes. We started our initial characterization with 11 mutants covering the different predicted functional categories (Table 3.2) and tested their ability to trigger invasome formation upon human endothelial cell infection. As reported in the research article III, this assay shows a stronger discriminative power to assess functionality of the VirB/D4 T4SS over measurements of the P\textsubscript{virB}:gfp reporter activity by flow cytometry or immunoblot. We included B. henselae wild-type and a ΔbatR deletion mutant as well as uninfected cells as controls and scored the percentage of invasome positive cells after 48 h of infection with a multiplicity of infection (MOI) of 20, 100 and 500. Infection with any of the tested transposon mutants resulted in a decrease of invasome positive cells compared to B. henselae wild-type infection (Fig. 3.10A), albeit to a different extent. At higher MOI (500 and 100), all tested transposon mutants were able to trigger invasome formation in at least few percent of the infected cells, in contrast to an infection with a ΔbatR mutant, which failed to induce invasome formation in any of the tested conditions. These results indicated that all the isolated transposon mutants still harbored some residual VirB/D4 T4SS activity. At a MOI of 20, more than 50% of the cells infected with wild-type bacteria still displayed a characteristic invasome structure, whereas infection with any of the transposon mutant showed a clear reduction, providing a functional confirmation for the attenuation of the VirB/D4 T4SS in these mutants. Clean deletion of the genes hit by
transposon insertion would be required to assess whether the observed residual VirB/D4 T4SS activity results from an incomplete inactivation of the disrupted gene.

![Graph](image)

**Figure 3.10: Invasome formation assay.** Quantification of invasome positive Eahy.926 cells after 48 h of infection at a MOI of 20,100 or 500. The bacterial strains used for infection are indicated. Mean value plus SD are represented for >300 cells from a representative experiment set-up in triplicate.

### 3.6.2. LC-MS analysis of *B. henselae* proteins expression in a subset of transposon mutants

We were interested in better understanding the underlying defect(s) resulting in an impaired induction of the *virB* promoter in the different isolated mutants. Therefore we assessed global changes in protein expression in total lysates of 11 individual transposon mutants after 48 h of host free pH-dependent induction using a shotgun proteomic approach combined with label free quantification of the detected proteins. The abundance of each detected protein was normalized to wild-type levels after 48 h of induction for each sample measured as technical duplicate. A total of 880 individual *B. henselae* proteins were identified in this analysis, corresponding to a coverage of 59% of the 1488 predicted *B. henselae* proteins. We focused our analysis on the proteins differentially regulated between the tested mutants and the wild-type strain after 48 h of M199/10% FCS induction. Proteins showing at least a 2 fold up- or down-regulation (q-value ≥ 0.05) were considered to be differentially regulated in our analysis.
We first investigated the proteins that were statistically differentially regulated in most of the investigated mutants compared to wild-type bacteria in order to identify any defect(s) common to these bacteria. From all the detected proteins, 39 showed co-regulation in at least 6 of the 11 transposon mutants with 30 down-regulated and 9 up-regulated proteins (Fig. 3.11 and Table 3.1). Not too surprisingly, these included all components of the VirB/D4 T4SS machinery detected in this experiment, i.e. the T4SS machinery components VirB5 and VirB11, the *Bartonella* effector proteins A, C, D and E, FicA, an homologue to the VbhA protein [18] encoded in front of BepA as well as the GFP expressed by the chromosomally encoded P<sub>virB-gfp</sub> reporter (Fig. 3.11A). These results confirmed the initial phenotype of the isolated mutants.

**Figure 3.11:** LC–MS analysis and label-free quantification of proteins in a subset of transposon mutants. Heat map representation of the proteins differentially regulated after 48 h of M199/10% induction. The criteria were ≥ 2-fold down (A) or up-regulation (B) in at least 6 of the 11 tested mutants with a qval ≤0.05). The protein expression was normalized to *B. henselae* wild-type after 48 h of induction. The name of the gene disrupted by transposon insertion is indicated above the column displaying the corresponding data. The left panel represents the quantification of the corresponding proteins in a ΔbatR or ΔrpoH1 mutant grown in the same condition (data from research article III).
Table 3.1: LC-MS Analysis of selected HimarI mutants affected in P\textsubscript{virB} induction

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>comM</td>
<td>Conserved exported protein</td>
</tr>
<tr>
<td>dprA</td>
<td>TASS protein V/VB</td>
</tr>
<tr>
<td>recA</td>
<td>Fragment of autotransporter</td>
</tr>
<tr>
<td>radA</td>
<td>Bartonella effector protein A</td>
</tr>
<tr>
<td>radC</td>
<td>Bartonella effector protein E</td>
</tr>
<tr>
<td>badA1</td>
<td>Transcriptional regulator</td>
</tr>
<tr>
<td>hbpC</td>
<td>Bartonella adhesin</td>
</tr>
<tr>
<td>batR</td>
<td>Green fluorescent protein</td>
</tr>
<tr>
<td>BH01910</td>
<td>Bartonella effector protein C</td>
</tr>
<tr>
<td>BH14070</td>
<td>Bartonella effector protein D</td>
</tr>
<tr>
<td>BH14150</td>
<td>Bartonella effector protein E</td>
</tr>
<tr>
<td>BH04980</td>
<td>Bartonella effector protein F</td>
</tr>
<tr>
<td>BatR</td>
<td>Bartonella effector protein G</td>
</tr>
</tbody>
</table>

Expression ratio of the LC-MS data presented in Fig. 3.11. The log2 ratio normalized to wild-type expression levels determined by label-free quantification are indicated. Ratios with an associated qval ≤ 0.05 are displayed in bold.
This prompted us to compare our list of differentially regulated proteins with those found to be affected in a batR mutant (research article III). Out of the 30 down-regulated proteins from our analysis, only 11 were also significantly down-regulated in a ΔbatR mutant, 8 of which were related to the VirB/D4 T4SS, and thus known targets of BatR. The remaining 19 proteins did not show any positive correlation with the proteins differentially regulated in a ΔbatR mutant, and neither did the 9 up-regulated proteins (Fig. 3.11A and Table 3.1). These results suggest the involvement of other(s) perturbation in the different mutants than the sole reduction of BatR expression level. Therefore we also compared our list of differentially regulated proteins with the one affected in a ΔrpoH1 mutant (research article III). Strikingly, 27 of the 30 proteins down-regulated in our analysis were also significantly down-regulated in a ΔrpoH1 mutant (Fig. 3.11.A and), suggesting that these mutants would be affected in an RpoH1-dependent process. None of the 9 down-regulated proteins were affected in a ΔrpoH1 mutant. We further investigated whether we could use the proportion of BatR and/or RpoH1 co-regulated proteins to distinguish individual mutants (Fig. 3.12). However, the trend was similar for each of the tested mutants, indicating a comparable contribution to the activation of this regulon. Yet, the proportion of RpoH1 regulated proteins was lower per mutant as in our global analysis, which likely reflects that other processes may be affected in these mutants.

**Figure 3.12: Contribution of BatR and RpoH1 to the differential regulation observed in a subset of transposon mutants.** The percentage of BatR (grey) or RpoH1 (black) regulated proteins among the proteins differentially down-regulated in each mutant are indicated. Note that a proportion of BatR-regulated proteins are also dependent on RpoH1 (see research article III).
3.6.3. Rescue of mutant phenotype by BatR or RpoH1 overexpression

The results of our LC-MS analysis showed a decreased level of BatR expression in all tested transposon mutants compared to wild-type bacteria, and revealed that most of the proteins down-regulated in these mutants were also affected in a ΔrpoH1 mutant. Unfortunately, we could not determine the expression level of RpoH1 in the different tested mutants as this protein was not detected in our analysis. To test whether the observed phenotype would be the result of a decreased level of one of these components, we assessed the effect of either BatR or RpoH1 overexpression in a subset of these mutants. Plasmids encoding an IPTG inducible copy of batR or rpoH1 were introduced in the mutant strains and the effect of over-expression on the induction of the VirB/D4 T4SS was assessed by flow-cytometry after 24 h of M199/10%FCS induction. Strikingly, either plasmid was able to rescue the induction of the virB promoter to almost wild-type levels in the presence of IPTG in each tested mutant (Fig. 3.13).

Figure 3.13: Induction of P\textsubscript{virB} is rescued by either BatR or RpoH1 overexpression. Expression of the chromosomal P\textsubscript{virB}/gfp reporter was determined is different B. henselae strains by flow-cytometry after 24 h of culture in M199/10%FCS in the presence (black bars) or absence (white bars) of 500 μM IPTG. Values represent the mean +/- SD of the percentage of GFP positive cells from three independent experiments compared to wild-type expression level (without plasmid). (A) The strains carrying a transposon insertion within the indicated gene and carrying the BatR overexpression plasmid pbatR are compared to wild-type bacteria without plasmid. (B) The strains carrying a transposon insertion within the indicated gene and carrying the RpoH1 overexpression plasmid prpoH1 are compared to wild-type bacteria without plasmid (n.d. : not determined).
These results support the hypothesis that the impaired P\textsubscript{virB} induction observed in the tested mutants results from a decrease in BatR and/or RpoH1 levels. Our epistasis analysis showed that although both BatR and RpoH1 are required for the expression of the VirB/D4 T4SS, RpoH1 is acting upstream from BatR \textit{(research article III)}. We would therefore conclude that all transposon mutants tested here are affected in processes located up-stream from RpoH1. As these results did not allow a clear distinction between the individual mutants tested, further experiments will be required to understand their relative contribution to the regulation of \textit{B. henselae} pathogenicity factors.

### 3.6.4. Results overview and possible roles of the targeted genes in P\textsubscript{virB} regulation

Our genetic screen aiming at identifying components involved in the regulation of \textit{B. henselae} pathogenicity in general and in the regulation of the \textit{virB} promoter in particular \textit{(research article III)} resulted in the isolation of a broad spectrum of mutants that were grouped into seven functional categories based on the predicted function of the disrupted gene. These categories were i) transcription regulation, ii) DNA repair and other DNA related activities, iii) diverse metabolic functions, iv) outer-membrane proteins, v) phage related proteins, vi) transport and vii) other unrelated processes. Most of the mutants assigned to the transcription regulation category have been characterized in the \textit{research articles I, II and III} and their contribution to \textit{B. henselae} host adaptation is relatively well understood. These included mutants in the histidine kinase BatS and in the stringent response components SpoT and DksA. In contrast, the contribution of the genes identified in the other isolated mutants in respect to the regulation of the VirB/D4 T4SS remained elusive. Here we present the preliminary characterization of a subset of these mutants (Table 3.2). Noteworthy, with the exception of the hemin binding protein C (HbpC), none of the identified genes or their encoded proteins had yet been studied in the bartonellacae. Their specific role in the biology of \textit{B. henselae} can thus only be elaborated based on studies conducted on orthologous proteins.

The quantification of invasome formation during endothelial cell infection as a read-out for VirB/D4 T4SS activity confirmed the primary phenotype of our selected subset of transposon mutants. LC-MS analysis of the proteome of these mutants verified reduced levels of the VirB/D4 T4SS. This analysis also revealed that its secreted effectors, the Bep proteins and one of their key transcription activator, the response regulator BatR were equally affected. Comparison with the protein expression pattern of a ΔbatR or a ΔrpoH1 mutant allowed us to place the individual defect of these mutants upstream from
the alternative sigma factor RpoH1. We further demonstrated partial complementation of some of these mutants by overexpression of either BatR or RpoH1, supporting the proposed epistasis. The fact that BatR overexpression appeared to rescue equally well the mutants as did RpoH1 is at first puzzling, considering the epistasis between RpoH1 and BatR. However, we have shown in the research article II that BatR overexpression results in a strong induction of RpoH1 expression. This suggests a reciprocal regulation for both factors, which could explain the convergent results presented here. As determined in research article III, the alternative sigma factor RpoH1 seems to be activated during the stringent response. The signal(s) and/or trigger(s) of the stringent response can vary a lot between different organisms, and even multiple triggers can be required for its activation [19]. However, these triggers have not yet been resolved for B. henselae. Nevertheless, our genetic data provide good evidences that the mounting of this response is affected in the different tested mutants. An interesting follow-up experiment would thus be to determine the cellular levels of the second messenger (p)ppGpp in the different isolated mutants. For those mutants where the defect in P_{virB} induction would correlate with decreased levels of the alarmone, rescue of the phenotype could be assessed by artificial induction of (p)ppGpp synthesis by ectopic expression of a truncated, metabolically active form of E. coli RelA [20].

**DNA repair and DNA integrity related mutants**

A prominent category of mutants isolated in the above mentioned screen comprises genes that have been described to be involved in (or associated to) the maintenance of DNA integrity in various bacteria. Here we selected 5 mutants from this category for further characterization. These different mutants carried a transposon insertion in comM (BH00440), dprA (BH08170), recA (BH10230), radA (BH05260) and radC (BH08340).

ComM (BH00440) shares homology to the *Haemophilus influenza* competence protein ComM (44% identity and same domain organization), an important factor for efficient incorporation of DNA into the chromosome after uptake by this naturally competent organism [21]. DprA (BH08170) shows homology to the DprA/Smf protein family [22]. These broadly distributed proteins have been shown to play an important role in chromosomal integration of DNA in bacteria showing natural transformation such as *H. influenza* [23] but also are encoded in many non-competent bacteria. For instance, DprA was shown to work as a single and double stranded DNA (ssDNA and dsDNA) binding protein in *E. coli* [24], where it was proposed to participate to RecA-mediated
DNA recombination [22]. RecA (BH10230, 70% identity with *E. coli* homologue) is a protein with highly pleiotropic functions that plays a central role in the repair and maintenance of DNA integrity in bacteria [25]. RecA binds to ssDNA to form a nucleoprotein filament that can perform a rapid and efficient search for homology in the double stranded chromosomal DNA and contributes to the strand exchange with the homologous template [26]. Besides its role in homologous recombination, RecA is also a key player in the induction of the SOS response, by contributing to the degradation of the repressor LexA upon binding to ssDNA [27]. RadA (BH05260) shares homology to the DNA repair protein RadA/Sms of *E. coli* (43% identity), which facilitates DNA repair after damages caused by UV radiation, X-rays, and chemical agents possibly by contributing to the resolution of Holliday junction [28]. RadA was also shown to repress the RecA-mediated SOS response during logarithmic growth of *E. coli*, presumably by reducing the half-life of ssDNA - a co-substrate required for the RecA mediated proteolysis of LexA - by favoring its repair [29]. Finally RadC (BH08340) is the homologue of *E. coli* RadC (35% identity), a protein that was first described to be involved in DNA repair. More recent reports have however shed much uncertainty about its role and its function remains largely uncharacterized [30].

The identification of so many mutants in genes that have been associated with DNA uptake in different naturally competent bacteria inevitably recalls the ancestral function of the VirB/D4 T4SS as a DNA transfer machinery [31]. In that context, an involvement of these factors in the regulation of the VirB/D4 T4SS could represent a regulatory control anterior to the evolvement of this T4SS as a protein effector secretion system. Moreover, conjugational DNA transfer is also a known inducer of the SOS response [32]. Considering the functional interaction between RecA and DprA [22] and knowing that *comM* is part of the SOS response regulon α-proteobacteria [27,33] it is tempting to assume that these associations are conserved in *B. henselae* and that they relate to the observed effect on the regulation of the *virB* promoter. A possible hypothesis is that the regulation of the *virB* operon is coupled to the activation of the SOS response, which may be triggered at the on-set of the host infection and could represent an additional signal used to activate host adaptation factors. The possible link between the SOS response and the activation of the VirB/D4 T4SS is reinforced considering that the second most important class of mutant isolated in our screen are affected in phage-related proteins (see below) and the SOS response is a well characterized inducer of such genes in many different bacteria [34,35]. Moreover, some reports have established a link between
the stringent response and the activation of the SOS response as in *E. coli*, where the SOS regulon is activated under stringent conditions [36]. Further experiments will be necessary to uncover the relation between the proteins affected in the different isolated mutants and how central processes such as the maintenance of DNA integrity or DNA damage signaling contribute to the modulation of host adaptation.

**Mutants in Bartonella phage related genes**

Prophages are known to play an important role in the genome dynamic of the bartonellacae. In each sequenced *Bartonella* genome, a variable number of prophages and phage related genomic islands have been described, with genes of phage origin scattered around the full genome of these bacteria [37–40]. Strikingly, a large proportion of the genes identified in our screen for mutants impaired in $P_{virB}$ induction carried an insertion in genes encoding phage-related proteins (*research article III*), most of which are encoded in the genomic region designed as BH-GI12 (spanning BH13900 to BH14090 [38]). This region encodes proteins that have been detected in purified bacteriophage-like particles (BLP) in several *Bartonella* species including in *B. henselae* [37] where they tend to accumulate in late logarithmic cultures [41]. BLPs have been proposed to play a role of gene transfer agent [37] although this property was never demonstrated under laboratory conditions. Interestingly, we have observed that the genes encoding the proteins found in BLPs are transiently up-regulated at an early stage of endothelial cell infection (*research article I*), and that this early response precedes the activation of the $virB/virD4/bep$ locus. Although a causal relationship between these two distinct processes remains to be demonstrated, a linkage between their expressions could represent a strategy to maintain a functional BLP locus in *B. henselae* throughout the infection cycle.

**Mutants in outer-membrane proteins and transporters**

The last category of mutants selected for this preliminary investigation carried a transposon insertion in genes encoding out-membrane proteins (OMPs) or an ABC transporter. The first OMP mutant carried an insertion in the gene coding for the *Bartonella* adhesin A1 (BadA1, BH01490), a short variant of the well characterized trimeric autotransporter BadA, which is an essential pathogenicity factor for the genus *Bartonella*, including *B. henselae* [42,43]. The second OMP mutant carried an insertion in the gene encoding the hemin binding protein C (HbpC, BH02550), which belongs to a conserved family of porin-like OMPs with hemin binding capacities (reviewed in [44]).
Strikingly, these two OMPs were also found among the top down-regulated proteins in all the mutants tested in our LC-MS analysis (Fig 3.10A and Table 3.1). These two proteins were also strongly down-regulated in an *rpoH1* mutant but up-regulated in a *batR* mutant compared to *B. henselae* wild-type. This could suggest the involvement of membrane homeostasis and/or membrane stress in the regulation of the VirB/D4 T4SS. This contribution could be rather indirect, as supported by the finding that a *B. henselae* mutant with an in-frame deletion of *badA1* (BH01490) did not show any defect in P\textsubscript{virB} induction, in contrast to the isolated transposon mutant (data not shown). This finding suggests that the transposon mutant phenotype resulted from some membrane perturbations, due for instance to the accumulation of aberrant and/or misfolded proteins in the periplasm of the bacteria.

**Further characterization**

This preliminary characterization of a subset of mutants affected in the regulation of the VirB/D4 T4SS leaves a lot of uncertainties regarding their specific contribution to the studied process. However, some patterns clearly emerge from our analysis. We have demonstrated that the stringent response constitutes an important signaling component for the activation of the VirB/D4 T4SS (*research article III*) and our analysis suggests that the defects in the different tested mutants would be upstream from this response. This signal integration over the stringent response could explain how apparently independent mutations result in a similar phenotype by converging towards the same stress response mechanism. Systematic generation of clean deletion mutants, epistasis analysis, and further phenotypic characterizations with a focus on stress response should allow clarify this picture. Although impressive, the coverage achieved by our shotgun LC-MS analysis was far from saturation. It varied between 59\% (3.4.2.1) and 68.7\% (*research article III*) of the proteins encoded in *B. henselae* genome, which may just not be sufficient to elucidate specific defect in any given mutant. Therefore additional read-outs with higher coverage (e.g. RNA-seq) would represent a valuable complementary approach. Further, the general characterization of the different stress responses in *B. henselae* would help to understand the observed interconnections between the different classes of mutants, and to gain a deeper understanding in the signal transduction networks that govern *B. henselae* adaptive response to its host environments.
3.7. EXPERIMENTAL PROCEDURES (related to unpublished results)

Bacterial strains, eukaryotic cell line, and growth conditions. *B. henselae* and *E. coli* strains were grown as previously described [3]. All cloning were performed in *E. coli* NovaBlue (Novagen). Plasmids were introduced into *B. henselae* by conjugation from *E. coli* ß2150 [45] using three-parental mating. Table 3.2 lists all strains and plasmids and Table 3.3 the sequence of all oligonucleotides used for the data presented in sections 3.5 and 3.6. The endothelial cell line Ea.hy926, resulting from a fusion of human umbilical vein endothelial cells (HUVEC) and the lung carcinoma cell line A549 [46] was cultured as reported previously [3].

Table 3.2: Strains and plasmids

<table>
<thead>
<tr>
<th>Strain or Plasmid</th>
<th>Relevant characteristic</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>E. coli</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Novablue</td>
<td>endA1 hsdR17r K12- m K12+ supE44 thi-1 recA1 gyrA96 relA1 lac(F' proA+B+ lacZAM15::Tn10 (Tet))</td>
<td>Novagen</td>
</tr>
<tr>
<td>ß2150</td>
<td>F' lacZAM15 lacY traD36 proA+B+ thrB1004 pro thi strA hsdS ΔdapA::erm (Ermr) pir</td>
<td>[4]</td>
</tr>
<tr>
<td><strong>B. henselae</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RSE247</td>
<td>Spontaneous SmR strain of ATCC 49882 T, serving as wild type</td>
<td>[3]</td>
</tr>
<tr>
<td>ABB127</td>
<td>ΔbatR in-frame deletion mutant of RSE247</td>
<td>This work</td>
</tr>
<tr>
<td>MQB435</td>
<td>ΔbatS in-frame deletion mutant of RSE247</td>
<td>This work</td>
</tr>
<tr>
<td>MQB242</td>
<td>ΔbatRS in-frame deletion mutant of RSE247</td>
<td>This work</td>
</tr>
<tr>
<td>MQB528</td>
<td>RSE247 carrying a chromosomal copy of PvirB:gfp</td>
<td>research article II</td>
</tr>
<tr>
<td>MQB279</td>
<td>RSE247 carrying a chromosomal copy of PbepD:gfp</td>
<td>This work</td>
</tr>
<tr>
<td>MQB529</td>
<td>ΔbatR deletion mutant carrying a chromosomal copy of PvirB:gfp</td>
<td>This work</td>
</tr>
<tr>
<td>MQB281</td>
<td>ΔbatR deletion mutant carrying a chromosomal copy of PbepD:gfp</td>
<td>This work</td>
</tr>
<tr>
<td>MQB531</td>
<td>ΔbatRS deletion mutant carrying a chromosomal copy of PvirB:gfp</td>
<td>This work</td>
</tr>
<tr>
<td>MQB277</td>
<td>ΔbatRS mutant carrying a chromosomal copy of PvirB:gfp</td>
<td>This work</td>
</tr>
<tr>
<td>MQT212</td>
<td>comM::Himar1 (BH00440, competence protein ComM)</td>
<td>research article III</td>
</tr>
<tr>
<td>MQT220</td>
<td>dprA::Himar1 (BH08170, DNA processing chain A)</td>
<td>research article III</td>
</tr>
<tr>
<td>MQT279</td>
<td>recA::Himar1 (BH10230, recombinase A)</td>
<td>research article III</td>
</tr>
<tr>
<td>MQT262</td>
<td>radA::Himar1 (BH05260, DNA repair protein RadA)</td>
<td>research article III</td>
</tr>
<tr>
<td>MQT317</td>
<td>radC::Himar1 (BH08340, DNA repair protein RadC)</td>
<td>research article III</td>
</tr>
<tr>
<td>MQT229</td>
<td>badA1::Himar1 (BH01490, adhesin)</td>
<td>research article III</td>
</tr>
<tr>
<td>MQT330</td>
<td>hbpC::Himar1 (BH02550, hemin binding protein C)</td>
<td>research article III</td>
</tr>
<tr>
<td>MQT266</td>
<td>BH01910::Himar1 (ABC transporter, ATP-binding protein)</td>
<td>research article III</td>
</tr>
<tr>
<td>MQT211</td>
<td>BH14070::Himar1 (phage-like protein)</td>
<td>research article III</td>
</tr>
<tr>
<td>MQT236</td>
<td>BH14150::Himar1 (hypothetical protein)</td>
<td>research article III</td>
</tr>
<tr>
<td>MQT210</td>
<td>BH04920::Himar1 (phage-like protein)</td>
<td>research article III</td>
</tr>
</tbody>
</table>
Table 3.2 (continued)

<table>
<thead>
<tr>
<th>Plasmids</th>
<th>Relevant characteristic</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>pCD366</td>
<td>gfp promoter probe</td>
<td>[13]</td>
</tr>
<tr>
<td>pDT024</td>
<td>batR under the control of Ptaclac for expression in Bartonella</td>
<td>[3]</td>
</tr>
<tr>
<td>pIT009</td>
<td>pPvwc:gfp; virB promoter region (bp -366 to +13) in pCD366</td>
<td>[3]</td>
</tr>
<tr>
<td>pIT011</td>
<td>pPvwc:gfp; bepD promoter region (bp -333 to +13) in pCD366</td>
<td>[3]</td>
</tr>
<tr>
<td>pMQ034</td>
<td>Derivative of pTR1000 used for deletion of batS</td>
<td>This work</td>
</tr>
<tr>
<td>pMQ009</td>
<td>Derivative of pTR1000 used for deletion of batRS</td>
<td>This work</td>
</tr>
<tr>
<td>pMQ016</td>
<td>Derivative of pIT009 carrying mutation 1 (prMQ1181/1182)</td>
<td>This work</td>
</tr>
<tr>
<td>pMQ021</td>
<td>Derivative of pIT009 carrying mutation 2 (prMQ1183/1184)</td>
<td>This work</td>
</tr>
<tr>
<td>pMQ022</td>
<td>Derivative of pIT009 carrying mutation 3 (prMQ1155/1156)</td>
<td>This work</td>
</tr>
<tr>
<td>pMQ023</td>
<td>Derivative of pIT009 carrying mutation 4 (prMQ1157/1158)</td>
<td>This work</td>
</tr>
<tr>
<td>pMQ017</td>
<td>Derivative of pIT009 carrying mutation 5 (prMQ1159/1160)</td>
<td>This work</td>
</tr>
<tr>
<td>pMQ018</td>
<td>Derivative of pIT009 carrying mutation 6 (prMQ1161/1162)</td>
<td>This work</td>
</tr>
<tr>
<td>pMQ024</td>
<td>Derivative of pIT009 carrying mutation 7 (prMQ1163/1164)</td>
<td>This work</td>
</tr>
<tr>
<td>pMQ025</td>
<td>Derivative of pIT009 carrying mutation 8 (prMQ1165/1165)</td>
<td>This work</td>
</tr>
<tr>
<td>pMQ026</td>
<td>Derivative of pIT009 carrying mutation 9 (prMQ1167/1168)</td>
<td>This work</td>
</tr>
<tr>
<td>pMQ027</td>
<td>Derivative of pIT009 carrying mutation 10 (prMQ1169/1170)</td>
<td>This work</td>
</tr>
<tr>
<td>pMQ028</td>
<td>Derivative of pIT009 carrying mutation 11 (prMQ1171/1172)</td>
<td>This work</td>
</tr>
<tr>
<td>pMQ019</td>
<td>Derivative of pIT009 carrying mutation 12 (prMQ1173/1174)</td>
<td>This work</td>
</tr>
<tr>
<td>pMQ020</td>
<td>Derivative of pIT009 carrying mutation 13 (prMQ1175/1176)</td>
<td>This work</td>
</tr>
<tr>
<td>pMQ029</td>
<td>Derivative of pIT009 carrying mutation 14 (prMQ1185/1186)</td>
<td>This work</td>
</tr>
<tr>
<td>pMQ030</td>
<td>Derivative of pIT009 carrying mutation 15 (prMQ1187/1188)</td>
<td>This work</td>
</tr>
<tr>
<td>pMQ033</td>
<td>To generate chromosomal insertion of Pvwc:gfp</td>
<td>This work</td>
</tr>
<tr>
<td>pMQ012</td>
<td>To generate chromosomal insertion of Pvwc:gfp</td>
<td>Research article II</td>
</tr>
<tr>
<td>pMQ085.3</td>
<td>pPvwc:dsRED; batR promoter region in pMQ366</td>
<td>This work</td>
</tr>
<tr>
<td>pMQ366</td>
<td>dsRED promoter probe, derivative of pCD366</td>
<td>This work</td>
</tr>
</tbody>
</table>
**Table 3.3: Oligonucleotides**

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cloning</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>prMQ1086</td>
<td>GCTCTAGATGAAACAGAAGGCTACCGAG</td>
<td>XbaI</td>
</tr>
<tr>
<td>prAB007</td>
<td>GCTCTAGATAGCAGCGCTCAATTCAGG</td>
<td>XbaI</td>
</tr>
<tr>
<td>prMQ1266</td>
<td>CTTTTTACTTTGGAAGGTCAGATTTGATCCTGCTTAATTTTC</td>
<td></td>
</tr>
<tr>
<td>prMQ1267</td>
<td>ACCCTTTACCAAAGTAGAAAGG</td>
<td></td>
</tr>
<tr>
<td>prMQ1088</td>
<td>TATCGAGCGAAGGGAAGGGCATTGGTTGGATTATGATCAT</td>
<td></td>
</tr>
<tr>
<td>prMQ1092</td>
<td>GCTCTAGATATATGGAAGGCGGCTTCGCTCAGATC</td>
<td></td>
</tr>
<tr>
<td>prGS28</td>
<td>GCCTACCCCTCCCACTG</td>
<td></td>
</tr>
<tr>
<td>prPE260</td>
<td>GTGTGACCAATGGCAGGAC</td>
<td></td>
</tr>
<tr>
<td>prMQ1181</td>
<td>GAGAAAGTTCTAACGAGGATGTGTTGC</td>
<td></td>
</tr>
<tr>
<td>prMQ1182</td>
<td>GCAACACATCCCTGTTAGAAACTTTCT</td>
<td></td>
</tr>
<tr>
<td>prMQ1183</td>
<td>GCCTTACCTAAGGCTGATGGTGC</td>
<td></td>
</tr>
<tr>
<td>prMQ1184</td>
<td>GCCTACATCAAGGCTAATTTTCTC</td>
<td></td>
</tr>
<tr>
<td>prMQ1185</td>
<td>GCTCTAGATGAAACAGAAGGCTACCGAG</td>
<td>XbaI</td>
</tr>
<tr>
<td>prMQ1186</td>
<td>CCTGCAGTATTTTGTTATTTGCAGAC</td>
<td></td>
</tr>
<tr>
<td>prMQ1187</td>
<td>CTTTTTACCAAAGTAGAAAGG</td>
<td></td>
</tr>
<tr>
<td><strong>EMSA competitor (forward sequence only)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>prMQ1217</td>
<td>TGCAATTAACCACTTATTTGCCG</td>
<td>Comp. 1 Fig. 3.4 prMQ1218</td>
</tr>
<tr>
<td>prMQ1219</td>
<td>CAGAGATTGTCGCTCAAATACCC</td>
<td>Comp. 2 Fig. 3.4 prMQ1220</td>
</tr>
<tr>
<td>prMQ1221</td>
<td>ACATTTTGCGAGGAACATTTTC</td>
<td>Comp. 3 Fig. 3.4 prMQ1224</td>
</tr>
<tr>
<td>prMQ1222</td>
<td>CTTTTTACTTTGAAGGAGG</td>
<td>Comp. 4 Fig. 3.4 prMQ1224</td>
</tr>
<tr>
<td>prMQ1223</td>
<td>AGAAATAATCCTTTTTTACAT</td>
<td>Comp. 5 Fig. 3.4 prMQ1226</td>
</tr>
<tr>
<td>prMQ1224</td>
<td>ATGTGTTGCAATTCACCTTACCAT</td>
<td>Comp. 6 Fig. 3.4 prMQ1228</td>
</tr>
<tr>
<td>prMQ1225</td>
<td>GCAAGATTTGCAATTTTACCATC</td>
<td>Comp. 7 Fig. 3.4 prMQ1228</td>
</tr>
<tr>
<td>prMQ1226</td>
<td>GCAAGATTTGCAATTTTACCATC</td>
<td>Comp. 8 Fig. 3.4 prMQ1228</td>
</tr>
<tr>
<td>prMQ1227</td>
<td>GCAAGATTTGCAATTTTACCATC</td>
<td>Comp. 9 Fig. 3.4 prMQ1228</td>
</tr>
<tr>
<td>prMQ1228</td>
<td>GCAAGATTTGCAATTTTACCATC</td>
<td>Comp. 10 Fig. 3.4 prMQ1228</td>
</tr>
<tr>
<td>prMQ1229</td>
<td>GCAAGATTTGCAATTTTACCATC</td>
<td>Comp. 11 Fig. 3.4 prMQ1228</td>
</tr>
<tr>
<td>prMQ1230</td>
<td>GCAAGATTTGCAATTTTACCATC</td>
<td>Comp. 12 Fig. 3.4 prMQ1228</td>
</tr>
<tr>
<td>prMQ1231</td>
<td>GCAAGATTTGCAATTTTACCATC</td>
<td>Comp. 13 Fig. 3.4 prMQ1228</td>
</tr>
<tr>
<td>prMQ1232</td>
<td>GCAAGATTTGCAATTTTACCATC</td>
<td>Comp. 14 Fig. 3.4 prMQ1228</td>
</tr>
<tr>
<td>prMQ1233</td>
<td>GCAAGATTTGCAATTTTACCATC</td>
<td>Comp. 15 Fig. 3.4 prMQ1228</td>
</tr>
<tr>
<td><strong>Probes for EMSA on PbatR</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>prMQ1237</td>
<td>GCTCTAGATGAAACAGAAGGCTACCGAG</td>
<td>PbatR- P1</td>
</tr>
<tr>
<td>prMQ1238</td>
<td>GACTTTTTAGCACCTTTCTATGC</td>
<td>PbatR- P1</td>
</tr>
<tr>
<td>prMQ1239</td>
<td>AACCAACACAAATTTTGAAC</td>
<td>PbatR- P2</td>
</tr>
<tr>
<td>prMQ1240</td>
<td>AACCTTTACGCTGATACGATCA</td>
<td>PbatR- P2</td>
</tr>
<tr>
<td>prMQ1241</td>
<td>CTTTTTACTTTGGAAGGTCAGATTTGATCCTGCTTAATTTTC</td>
<td></td>
</tr>
<tr>
<td>prMQ1242</td>
<td>AAAATTTGCTGTTTTCTGCATC</td>
<td>PbatR- P3</td>
</tr>
<tr>
<td>prMQ1149</td>
<td>ATGGAAAGTGCTTCAACACATAC</td>
<td>PbatR- P4</td>
</tr>
<tr>
<td>prMQ1150</td>
<td>GTGTTAACCCTTTGAGAGAGATC</td>
<td>PbatR- P4</td>
</tr>
</tbody>
</table>
Plasmid Constructions

Construction of plasmids for the generation of in-frame deletions in B. henselae.
In-frame deletion mutants were generated by a two-step gene replacement procedure as described [47, 48]. All plasmids used for in-frame deletion were constructed in the background of pTR1000 [48]. The oligonucleotides used for the construction of these plasmids are listed in table 3.2.

Construction of mutated GFP reporter plasmids. The derivatives of pP$^\text{virB-gfp}$ harboring mutation in the BatR-binding motif were created by megaprime PCR using the oligonucleotides designed for EMSA competition assay (Table 3.3) together with prGS28 (fragment 1) and prPE260 (fragment 2) using pP$^\text{virB-gfp}$ as template. The two fragments were combined using prGS28 and prPE260, purified, digested by BamHI/EcoRI and cloned into the corresponding sites of pCD366. For the mutations 1, 2 and 14, 15 (pMQ016, pMQ021, pMQ029 and pMQ030), new primers were designed to incorporate the mutations (Table 3.2 and 3.3), as their distal location on the oligonucleotides resulted in a wild-type promoter after combination of both fragments. All mutations were verified by sequencing of the resulting plasmids.

Infection Assays. Infections of Ea.hy926 cells were performed as previously described [3]. In brief, cells were grown to confluency in Dulbecco’s Modified Eagle Medium (DMEM with Glutamax, Gibco Invitrogen) supplemented with 10% fetal calf serum (FCS, Gibco Invitrogen) in a humidified atmosphere at 37°C and 5% CO$_2$. One hour before infection, the cells were washed with Medium 199 (M199, Gibco Invitrogen) supplemented with 10% FCS. The bacteria were grown on CBA plates for 48 h, harvested in M199/10%FCS (pH 7.4) and used to infect the Ea.hy926 cells at a multiplicity of infection of 200, unless stated differently. The infected cells were incubated in a humidified atmosphere at 35°C and 5% CO$_2$ for 48 h.

Flow Cytometry. B. henselae strains were grown on Columbia agar plates containing 5% defibrinated sheep blood (CBA plates) supplemented with appropriated antibiotics and grown in a humidified atmosphere at 35°C and 5% CO$_2$ for three days followed by re-streaking on fresh CBA plates and growth for 48 hours. The bacteria were
Results - Experimental procedures related to unpublished results

resuspended in M199/10%FCS at a final OD$_{600}$ nm of 0.008 and incubated in 48 well plates in a humidified atmosphere at 35°C and 5% CO$_2$. Expression of the P$_{virB\_gfp}$ promoter was measured as GFP fluorescence by using a FACSCalibur flow cytometer (Becton Dickinson) with an excitation at 488 nm. Data analysis was performed using the FlowJo.

**Electrophoretic mobility shifts assays (EMSA).** Assays were performed as previously described [3]. In short, radiolabeled probes were generated by PCR reaction in presence of [$\gamma$$^3$P]ATP using prIT011 and prDT015 and purified pP$_{virB\_gfp}$ as template. Reactions were purified using a nucleotide removal kit (Qiagen). For each probe preparation, a parallel reaction was performed in absence of radioactive dATP, and DNA concentration was determined using a NanoDrop ND-100 spectrophotometer (Thermo Scientific). Binding reactions were performed in a total volume of 20 μl of buffer A (50 mM Tris [pH8.5], 300 mM NaCl, 2 mM DTT and 10% glycerol) in presence of 1 μg of poly(dI:dC) and 2-4 fmol radiolabeled probe. The reaction mixtures were separated on 5 to 8% polyacrylamid gel in 0.5% Tris/borate/EDTA buffer at 120 V for 1 h after what the gels were dried. Radioactivity was quantified with a Molecular Dynamics Typhoon 8600 phosphorimager. For competition assays with non-labeled double stranded oligonucleotides, the reaction was supplemented with a 1000 molar excess of annealed competitors.

**Host cell-free induction of B. henselae for protein analysis:** Bacteria were grown on CBA plates for 48 h, harvested in M199/10%FCS (pH 7.4), washed once and diluted in 25 ml M199/10%FCS to a final OD$_{600}$ of 0.065/ml. Bacteria were incubated in 150 cm$^2$ cell culture flasks for 48 h at 35°C and 5% CO$_2$ in a humidified atmosphere, harvested by centrifugation 5 min at 4,800 X g in a swinging-bucket and resuspended in M199/10%FCS. For mass spectrometry analysis, 1 ml of each bacterial suspension (OD$_{600}$ = 1) were centrifuged, the pellets were snap frozen in liquid nitrogen and stored at -80°C until further processing. For immunoblot analysis, the bacterial suspension was resuspended in 1xPBS to an OD$_{600}$ of 16/ml, mixed with 1 volume 2xSDS sample buffer, heated for 5 min at 95°C and stored at -20°C until further processing.

**Immunoblot analysis.** Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and immunoblotting for the detection of BatR and VirB5 and BepD proteins
Results - Experimental procedures related to unpublished results

were performed as described [3]. In short, B. henselae cells were harvested after 48 h growth in M199/10%FCS, washed 1x in PBS, resuspended to an OD_{600} of 16 and mixed with an equal volume of 2X Laemmli buffer. For each sample, 10 µl were separated by 14% SDS-PAGE and transferred to a nitrocellulose membrane (Hybond-C Extra, GE Healthcare). The immunoblot was developed with polyclonal rabbit sera raised against recombinant BatR (1:20'000) or VirB5 (1:50'000), followed by a 1:15’000 dilution of a goat anti-rabbit horseradish peroxidase-conjugated secondary antibody (GE Healthcare). Immunoblots were developed using LumiGLO chemiluminescent substrate (KPL) and imaged using an ImageQuan LAS 4000 (GE Healthcare).

Analysis of B. henselae proteins by LC-MS/MS. The sample preparation was performed as described in research article III. In brief, pellets were digested in 50 µl lysis buffer (8M urea, 0.1% RapiGest, 0.1M ammoniumbicarbonate), disrupted by sonication (Hielscher Ultrasonicator) and protein concentration was determined by BCA assay (Thermo Fisher Scientific). Lysates were reduced with 5 mM TCEP for 60 min at 37°C, alkylated with 10 mM iodoacetamide for 30 min in the dark and quenched with 12.5 mM N-acetyl-cysteine. Proteins were digested by addition of Lys-C (Wako) for 4 h at 37°C (protein to Lys-C ration 100:1), diluted to a final urea concentration of 1.5 M with 100 mM ammonium bicarbonate buffer and further digested by addition of Trypsin (Promega) and incubation at 37°C for more than 15 h (protein to Trypsin ratio: 50:1). After digestion, the samples were supplemented with TFA to a final concentration of 0.5% and HCl to a final concentration of 50 mM. Peptides were desalted on C18 reversed phase spin columns according to the manufacturer’s instructions (Microspin, Harvard Apparatus), dried under vacuum and stored at -80°C until further processing. LC–MS/MS analysis of digested and purified B. henselae lysates was performed on a dual pressure LTQ-Orbitrap mass spectrometer (Thermo Electron) connected to an electrospray ion source (Proxeon Biosystems) as described (research article III and [49]). Label free quantification was performed using Progenesis software (Nonlinear Dynamics, Version 4.0). The peptide false discovery rate (FDR) was set to 1% on the peptide level and validated using the number of reverse protein sequence hits in the data sets. Peptide abundance ratios (integrated peak areas) across different digestion conditions and accompanying q-values (i.e., p-values adjusted for multiple testing) were calculated as described using SafeQuant module and the QVALUE R package [49].
3.8. REFERENCES


4. Concluding remarks
CONCLUDING REMARKS

We have investigated the transcription mechanisms that control the expression of the host adaptation factors in the zoonotic pathogen *Bartonella henselae*. We describe the changes in gene expression taking place in conditions that mimic the infection of the mammalian host. We show that this dynamic process is associated with extensive remodeling of the bacterial membrane composition, with the differential regulation of *Bartonella* specific autotransporters, adhesins, hemin binding proteins, and T4SSs. We were especially interested in the regulatory control of the VirB/D4 T4SS and its secreted Bep effectors. Not only do these factors represent essential host adaptability components for the genus *Bartonella* [1], they also likely contributed to the spectacular adaptive radiation observed for these bacteria [2]. Bacteria of the genus *Bartonella* are indeed remarkable by their stealthy infection strategy, that enables these α-proteobacteria to cause long lasting and asymptomatic infection in their mammalian reservoir host [3].

*Sensing the environment*

We have shown that *B. henselae* host adaptation strategy relies on BatR/BatS, a vertically inherited regulatory system that was adopted to control the regulation of *Bartonella* specific genes, including these encoding the VirB/D4 T4SS and its secreted effector proteins (*Research article I* [4]). Noteworthy, there is only a very limited overlap between the regulons of orthologous regulators in closely related pathogenic or symbiotic bacteria [5,6]. This supports a role for transcriptional rewiring of the BatR/BatS regulated targets in the process of host adaptation [7]. We also propose that the switch of signal specificity from acidic- to physiological pH by the HK BatS was an important adaptation step in *B. henselae* infection strategy. This newly evolved signal specificity allows the distinction between the mammalian host environment, where the expression of BatR-regulated genes are required, and the midgut of the arthropod vector, which likely requires an alternative set of adaptation factors [4]. The genetic and regulatory requirements for *B. henselae* to persist in its arthropod host have not yet been comprehensively addressed and thus constitute an attractive topic for further research.

The molecular mechanism of signal perception by the HK BatS have not been resolved to date, and whether BatS directly senses environmental pH remains to be demonstrated. Recently, a short periplasmic protein termed ExoR was shown to modulates the activity of the BatS ortholog in *Sinorhizobium meliloti* and *Agrobacterium*
Concluding remarks

*Concluding remarks.* ExoR was shown to interact with the periplasmic domain of BatS and to repress its activity [10,11]. In *A. tumefaciens*, this accessory was further shown to undergo proteolytic degradation in conditions where the ortholog of BatS is activated. This finding strongly suggest a direct role for ExoR in the signal perception by BatR/BatS related TCSs [11]. Characterization of *B. henselae* ExoR homologue may thus reveal new insights into the mechanism of host recognition by the BatR/BatS TCS.

**Stringent response and host adaptation**

We have discussed how direct perception of the environment is critical for host adaptation. In this work, we also show that this process is intimately linked to the physiological state of the bacteria. Indeed, our results support a central role for the SR components SpoT and DksA in the expression of *B. henselae* host adaptation repertoire throughout its infection cycle (*research article III*). We further propose that the SR signaling has been adopted for the fine tuning of the mammalian host infection, as exemplified by the differential regulation of *B. henselae* VirB/D4 and Trw T4SSs by SpoT and DksA *in vitro*. We have also good supporting evidences for a similar role *in vivo*, as SpoT was identified as essential for *B. tribocorum* to colonize its natural host, the rat [1]. Although the natural host of *B. henselae* can be infected under laboratory conditions (e.g. [12]), we did not use the cat infection model in this work. Rather, we think that more detailed studies related to the SR control of *Bartonella* host adaptation *in vivo* should be addressed with surrogate models. The related *Bartonella* species *B. tribocorum* and *B. birtlesii* would be good candidates as the infection of their respective host (the rat and the mice) are well established. Of primary interest would be the *ex-vivo* characterization of bacteria recovered from the intraerythrocytic niche, using global methods such as RNA-seq or shotgun proteomics. The lack of established *in vitro* conditions to study the pathogenicity of either *B. tribocorum* or *B. birtlesii* supports the further use of *B. henselae* as an *in vitro* model for detailed studies on the molecular basis of host adaptation in the genus *Bartonella*.

We describe the mechanism by which the SR controls the expression of the VirB/D4 T4SS and its secreted effector proteins, via the recruitment of the alternative sigma factor RpoH1. Yet much uncertainty remains about the triggers of the SR in *B. henselae*. Although SR components are found in almost every bacteria with a high degree of conservation [14], the activating triggers or signals appear to be very diverse between different organisms. Similarly, the specific response to a SR activation widely
varies between bacteria [15,16]. This indicates that this central regulatory mechanism has been independently adopted to control the adaptation to many different life styles and habitats. Amino acid starvation contributes to the induction of the SR in many bacteria. For instance, single amino acid starvation is sufficient to trigger the SR in \textit{E. coli} by activation of the (p)ppGpp synthetase RelA [17]. In contrast, bacteria encoding a single bifunctional RelA-SpoT homologue (RSH) may require several starvation inputs to activate the SR. For example, SR activation in \textit{Caulobacter crescentus} require both amino acids and either carbon or nitrogen starvation [18]. It was shown that \textit{B. henselae} does not catabolize glucose but rather amino acids [19]. It is thus likely that amino acids availability contributes to the activation of the SR in this organism. Whether other metabolic signals are involved in the mounting of this response remains to be elucidated. However, determination of the specific triggers activating \textit{B. henselae} SR \textit{in vitro} would help to understand the physiological relationship between the regulations patterns we observed and the microenvironments encountered within the host.

Our screen for \textit{B. henselae} mutants affected in expression of the VirB/D4 T4SS unexpectedly resulted in the identification of many apparently unrelated mutations. Characterization of a subset of these mutants revealed very similar protein expression alterations as those observed in mutants directly affected in SR signaling (\textit{research article III} and \textit{unpublished results}, section 3.6). Moreover, a small scale epistasis analysis indicates that these mutations act upstream from RpoH1, as overexpression of this factor rescued the mutants phenotype. This suggests that the SR may represent an essential signaling hub for \textit{B. henselae}, which would also integrate non-metabolic inputs. Considering the absence of RpoS homologue in α-proteobacteria [20], we can hypothesize that the SR machinery has been adopted by these bacteria to coordinate their general stress response. Further studies are indeed required to understand the interconnections between the different classes of isolated mutants, and to gain a deeper understanding in the signal transduction networks that govern \textit{B. henselae} adaptive response to its host environments.

\textit{Convergent promoter architecture}

Our preliminary characterization of the vertically inherited \textit{batR} promoter has revealed unexpected similarities to the horizontally acquired \textit{virB} promoter (\textit{research article I, III} and section 3.4). First, both promoter regions depend on the dual input of BatR and on the $\sigma^{32}$ RpoH1 for their transcription. Further, both promoter regions display a BatR-independent transcription activity in the absence of the response regulator and
finally, both promoters exhibit a bistable behavior under activating conditions. The finding that the *batR* promoter is only partially dependent on its own product is perfectly in line with the situation described for canonical TCS. Their expression typically depends on two promoters, a weak constitutive promoter required for the basal expression of the TCS and a second promoter, positively controlled by the activated RR, which constitutes the positive feed-back loop [21,22]. The biological relevance of the BatR-independent transcription observed at *P_{virB}* remains uncertain. It is tempting to speculate that this transcription activity represents the ancestral regulatory state of this horizontally acquired locus. Interestingly, this activity is not observed in conditions where BatR is inactive. It is only seen in strains where the *batR* gene has been deleted. This suggests that this BatR-independent activity is repressed by BatR or by a BatR-regulated factor. An interesting candidate for further investigations of the regulation of these promoters is RosAR (BH04610), a DNA binding of the Ros/MucR family which was down-regulated in BatR and RpoH1 regulons and which is known to act as transcriptional repressor in α-proteobacteria [23].

The inherent bistability observed for *batR* and *virB* promoters is an interesting finding and we can imagine that the resulting population heterogeneity can contributes to *B. henselae* host adaptation strategy. A straightforward approach to verify this hypothesis would be the engineering monostable variants of these promoters in their native chromosomal context, using *B. tribocorum* or *B. birtlesii* as model. Assessing the infectivity of such monostable variant compared to their parental strain would allow quantifying the contribution of bistable expression of the VirB/D4 T4SS for host colonization. A similar approach was used to demonstrated the importance of the positive feedback regulation of PhoP/PhoQ TCS *Salmonella enterica* in the mice infection model [22,24]. The source of the bistability observed for *P_{batR}* and *P_{virB}* transcription has not yet been unambiguously resolved. It is conceivable that this behavior originates from the positive feedback regulation of the BatR/BatS TCS. Although prototypical TCSs are not prone to bistability [25], several TCSs with non-canonical chromosomal organization, such as genetic uncoupling of both RR and HK, have been shown to trigger bistable expression [26,27]. Whether the genetic organization of the *batR* *batS* locus, with the presence of a short intergenic region between both genes and a *batS* promoter within the BatR coding sequence would be sufficient to sustain bistable behavior constitutes an attractive hypothesis that remains to be demonstrated. Alternatively, the observed bistability could reflect RpoH1 contribution to *P_{batR}* and *P_{virB}* activity. Although our data
suggest that (p)ppGpp is not involved in the bistable property these promoters, it could result from post-translational control of RpoH1 at the single cell level.

References


5. Acknowledgments
ACKNOWLEDGMENTS

This work was carried out in the group of Prof. Christoph Dehio in the Focal Area of Infection Biology at the Biozentrum of the University of Basel.

First of all, I would like to thank Christoph Dehio for supporting my decision to start my PhD in his research group, after these years of work as a technical associate. I am grateful to his trust, his support and his guidance, as well as to the freedom given to explore the mysteries of *Bartonella* transcription regulation.

I would also like to thank Prof. Urs Jenal and Prof. Dirk Bumann, who accepted to take part to my thesis committee. Their advices and feedbacks have been very valuable to my research.

Many thanks to all colleagues with whom I had the chance to work with in the “485-microcosme”. Dr. Michaela Dehio, who introduced me to the genome wide aspects of *Bartonella*; Andrea Basler for her enthusiasm; Dr. David Tropel, to raise my awareness on the world of sigma factors; Dr. Gunnar Schröder, for his friendship; Dr. Philipp Engel, for great scientific discussions, memorable “key note lectures” as well as for sharing his enthusiasm; Marius Liesch, for his cheerfulness; Dr. Arnaud Goepfert, guest of 485, to make me aware of structural biology and for nice discussions.

Special thanks to Alexander Harms, for his enthusiasm and for sharing his endless scientific curiosity, as well as for critical reading my manuscript.

Another special thanks to Mathias Dick and his decision to join the project. Besides his great contribution to *Bartonella* research, it has been a real pleasure working with him.

I want to thank all current members of the laboratory and the institute who contribute to this very rich working environment and who helped me one or the other way. I would specially like to thank Dr. Simone Eicher for her detailed comments on my manuscript. I would also like to thank Dr. Houchaima Ben Tekaya, Alain Casanova, Shyan Low, Yun-Yueh Lu, Simone Muntwiller, Dr. Rusudan Okujava, Kathin Pieles, Frédéric Stanger, and Sarah Stiegeler. I thank Claudia Mistl for her help. I am also grateful to Claudia Erbel-Sieler, Michaela Hanisch, Marina Kuhn-Rüfenacht, Margot Richter-Auer.
and Roger Sauder for their assistance over the past years. I would also like to thank Dr. Samuel Steiner, Dr. Jake Malone and Dr. Alain Mazé for nice discussions. Thanks to Mario Emmenlauer and to Dr. Pauli Rämö for their computational insights. I would also like to thank all former members of the laboratory, especially Dr. Ralf Schülein and Dr. Anja Seubert, Christopher Snyder, Isabella Toller, Dr. Arto Pulliainnen, Marco Faustmann, Matthias Truttmann, Patrick Guy, Jérémie Gay-Fraret, Florine Scheidegger and Dr. Raquel Conde. And a big thank to those I forgot to mention, sorry for that.

Merci to Dr. Thomas Rhomberg, Dr. Stephan Jungblut and Dr. Thomas Walpen, who played an important role in my decision to start this journey. Merci beaucoup à Dr. Christophe Bodenreider for his support.

Last but not least, I want to thank Dr. Gabriela Klocek for her constant support over the last years, for sharing the up-and-downs and so much more. I don’t think I would have made it through without your love. Dziękuję bardzo moje kochanie, jesteś fantastyczna.
6. Curriculum Vitae
CURRICULUM VITAE

Last Name: Québatte
First Name: Maxime
Date of Birth: 09.12.1974
Citizenship: Swiss

Office Address: Klingelbergstrasse 50/70
               CH-4056 Basel, Switzerland
               +41 61 267 2138
               m.quebatte@unibas.ch

Private address
               Solothurnerstrasse 92
               CH-4053 Basel, Switzerland
               +41 78 886 86 53

EDUCATION AND RESEARCH EXPERIENCE

2009 - present  PhD Thesis
                 Focal Area Infection Biology, Biozentrum, University of Basel, CH
                 Supervisor: Prof. Christoph Dehio
                 Committee members: Prof. Urs Jenal and Prof. Dirk Bumann,
                 Title: Uncovering the transcriptional control of Bartonella henselae
                 host adaptation factors

2001 - 2009  Technical Associate
              Focal Area Infection Biology, Biozentrum, University of Basel, CH
              Group of Prof. Christoph Dehio
              Main topic: Development of DNA spotted microarray,
              Gene expression analysis using microarray and quantitative RT-PCR.

1999 - 2000  Research Assistant
              Friedrich Miescher-Institut, Basel, CH
              Group of PD Dr. Jerzy Paszkowski
              Topic: "Homologous recombination and gene silencing in Arabidopsis thaliana"

1997 - 1998  Diploma Thesis
              University of Neuchâtel, Biochemistry, CH
              Group of Pr. Jean-Marc Neuhaus
              Title: "Vacuolar sorting of proteins in the moss Physcomitrella patens"

1993 - 1998  Undergraduate studies:
              Diploma in Biology, University of Neuchâtel, CH
              Specialization: Biochemistry and Plant Physiology

1990-1993  Matura type C, Gymnase de la Chaux-de-Fonds, CH
TEACHING EXPERIENCE

2012-2013  Supervision of a master student
Focal area Infection Biology, Biozentrum, University of Basel, CH

2001-2003  Co-supervision of master students
Focal area Infection Biology, Biozentrum, University of Basel, CH

2001 - present  Assistant and tutor for undergraduate students in practical courses
“Microbiology”
Focal area Infection Biology, Biozentrum, University of Basel, CH

1995-1997  Assistant for undergraduate students in practical courses in plant
physiology and biochemistry
University of Neuchâtel, CH

ORAL AND POSTER PRESENTATIONS

2013  Poster presentation at the ESF conference Bacterial Networks 2013, Pułtusk, Poland. Title: “Dual signal input: regulation of B. henselae VirB/D4 Type IV secretion system by the BatR/BatS two component system and the alternative sigma factor RpoH1”

2012  Poster presentation at Infection Biology Symposium, Biozentrum, Basel, CH. Title: “Regulation of Bartonella henselae Type IV Secretion System During Human Endothelial Cell Infection”

2012  Poster presentation and first prize at the BioValley Life Science Week, Basel, CH. Title: “Regulation of Bartonella henselae type IV secretion system during human endothelial cell infection”

2010  Poster presentation at the ESF conference Bacterial Networks 2010, Sant Feliu de Guixols, Spain. Title: “Regulation of Bartonella henselae adaptive response during endothelial cell infection”

2010  Oral presentation at Presentation at the Swiss Molecular Microbiology Conference (SWIMM10), Interlaken, CH. “The BatR/BatS two component regulatory system controls the adaptive response of Bartonella henselae during human endothelial cell infection”
PUBLICATIONS


