CHAPTER 4

SYMBIOSIS: VARIATION IN *BUCHNERA APHIDICOLA*’S LEUCINE PLASMID MAY CONFER ADVANTAGE TO RUSSIAN WHEAT APHID BIOTYPE
Introduction

Aphids’ success as pests can be attributed to a symbiotic alliance with *Buchnera aphidicola*, a bacterium that allows them to exploit dietary imbalanced phloem sap as food source (Srivastava 1987; Munson *et al.* 1991; Douglas 1998; Blackman & Eastop 2000). This ancient relationship between aphid and symbiont is crucial for the survival of both organisms, with the elimination leading to death of its partner (Munson *et al.* 1991; Lai *et al.* 1994; Douglas 1998; Sandstrom & Moran 1999; Baumann *et al.* 2006). Indeed, this relationship where the symbiont is located within host produced cells (bacteriocytes/ mycetocytes) are regarded as an advanced stage of symbiosis (Dixon 1998; Douglas 1998; Braendle *et al.* 2003). Here the bacterial symbiont is responsible for the synthesis and recycling of specific essential amino acids that is either present at low concentrations or absent in the dietary phloem (Douglas & Prosser 1992; Prosser & Douglas 1992; Febvay *et al.* 1994; Lai *et al.* 1994; Douglas 1998; Thao *et al.* 1998; Sandstrom & Moran 1999; Baumann *et al.* 2006).

*Diuraphis noxia* (Russian wheat aphid, RWA) is a major pest found in all but a few of the cereal producing countries (Blackman & Eastop 2000). The RWA only contains *B. aphidicola* as its endosymbiont (Swanevelder *et al.* 2010). However, in this relationship the contribution of the endosymbiont in the maintenance of certain essential amino acids is questionable when compared to other cereal feeding aphids. Lower gene copy numbers and the presence of pseudogenes (Lai *et al.* 1996; Wernegreen & Moran 2000; 2001), together with lower plasmid copy numbers (Baumann *et al.* 1995; Lai *et al.* 1996; Rouhbakhsh *et al.* 1996; Silva *et al.* 1998; Thao *et al.* 1998; Baumann *et al.* 1999; Soler *et al.* 2000) and higher non-synonymous substitutions rates in functional amino acid biosynthetic genes (Wernegreen & Moran 2000), all suggest a reduced contribution towards essential amino acid biosynthesis by the endosymbiont and/or a lower dependency of the aphid host on its symbiotic partner. This degradation in the mutualistic relationship is attributed to the ability of the RWA to induce higher levels of the selected essential amino acids in the phloem sap of susceptible host plants, thereby removing the selective pressure from the endosymbiont to retain as many functional gene copies of the required essential amino acid biosynthetic genes (Telang *et al.* 1999; Porter & Webster 2000; Sandstrom *et al.* 2000; Ni *et al.* 2001).

The recent appearance of new RWA biotypes in the USA and South Africa (Haley *et al.* 2004; Burd *et al.* 2006; Tolmay *et al.* 2007; Weiland *et al.* 2008) allowed for the investigation of aphid biotype variation and development in the field (Lapitan *et al.* 2007; Shufran *et al.* 2007). RWA biotypes are not anatomically or morphologically distinguishable from each other, but are discernable based on their ability to overcome specific host resistances in a plant differential
study (Puterka et al. 1992; Jyoti & Michaud 2005; Burd et al. 2006; Jyoti et al. 2006; Weiland et al. 2008). Furthermore, both symbiotic partners showed little sequence variation between different biotypes (Lapitan et al. 2007; Shufran et al. 2007; Swanevelder et al. 2010). However, in the aphid-endo symbiont relationship, a small change in a symbiont could have dire consequences for the aphid host, e.g. a single point mutation in Buchnera can determine aphid heat tolerance (Dunbar et al. 2007).

In a previous study on RWA B. aphidicola, the only sequence variation identified from the different D. noxia biotypes was a CCC-insert located between the repA2 and leuA genes of the leucine plasmid (Swanevelder et al. 2010). The insert, though in a non-coding region, not only introduces a new predicted rpoH3 (σ32) transcription factor binding site (TFBs) to the leucine plasmid (Swanevelder et al. 2010), but is also located within an Aphididae conserved inverted repeat region. This may suggest some functional constraints for the genome (Silva et al. 1998). Furthermore, the newly predicted rpoH is one of only two sigma TFs predicted from sequenced B. aphidicola genomes, i.e. σ32 and σ70 (Shigenobu et al. 2000). To date only a single regulatory gene involved in essential amino acid regulation was identified from the many Buchnera genomes sequenced. Also, B. aphidicola only has one regulatory pathway that could regulate essential amino acid biosynthesis (Moran et al. 2005). It is thus plausible that transcriptional regulation via plasmid copy number is the most likely regulatory mechanism.

Variations in leucine plasmid copy number was found for Buchnera from different RWA biotypes (Moran et al. 2003; Swanevelder et al. 2010). The location of the CCC-insert, i.e. in the conserved inverted repeat and near predicted promoters, and the newly introduced TFBs, then suggest that this CCC-insert may play a role during transcription regulation of the leucine plasmid genes. Here we investigate the functionality of the CCC-insert different RWA biotypes. We also examine the sequence variation of the inverted repeat region located between the repA2 and leuA of the leucine plasmid as an indicator of the potential regulatory utilization within other species of the family Aphididae.

Materials and methods

Aphids

The original South African D. noxia biotype (SA) was obtained from the ARC-Small Grains Institute, Bethlehem, South Africa, and maintained on a susceptible wheat cultivar Scheepers. A mutated form of the SA biotype (SAM) was maintained on a resistant wheat cultivar TugelaDN (Van Zyl & Botha 2008). Females of both South African biotypes were kept
in insect cages at 20 ± 2 °C with continuous artificial fluorescent lighting. The USA biotypes were obtained from Prof. N. Lapitan (Colorado State University, Fort Collins, USA) where they were maintained on a mixed diet of susceptible wheat and barley cultivars under greenhouse conditions (Lapitan et al. 2007). Samples of Diuraphis mexicana and D. tritici, collected in Colorado, were kindly provided by Dr. G.J. Puterka (USDA-ARS, Stillwater, Oklahoma, USA) and specimens of Brevicoryne brassicae (MF1435), Hyalopterus pruni (MF1422), Macrosiphum rosae (MF1408), Myzus persicae and Uroleucon sonchi (ACAM937) were a donation from Mr. I. Millar (South African National Collection of Insects, PPRI, Pretoria, South Africa).

**DNA and RNA extraction**

Aphid total DNA was extracted using the DNAzol extraction protocol (Molecular Research Centre, Cincinnati, USA) and cleaned with the DNeasy cleanup kit (Qiagen, USA) that included the on column RNase treatment (Qiagen). All samples were quantified using the Nanodrop ND-1000 Spectrophotometer (Thermo Scientific, RSA). RNA extractions were performed on a 100 individuals collected with a soft brush and immediately frozen with liquid nitrogen. All extractions were done in accordance to the manufactures’ protocols using the RNeasy Mini Kit (Qiagen) which included the on column DNase I treatment (RNase-free DNase set, Qiagen), before spectrophotometric quantification.

**Leader sequence determination**

The leader sequences for Buchnera of the different RWA biotypes, with the insert and without the CCC-insert, were determined using the 5’ RACE (rapid amplification of cDNA ends) system (Version2.0E, Invitrogen, USA). Two reverse primers were designed from the Genbank accession FJ705299, LeuA_RACE1_R (5’-CATTGCATCACCTGCTACCT-3’) 244 bp into the leuA gene, and a nested primer (LeuA_RACE2_R, 5’-GTAATGCTTGTTCACCATC-3’) 36 bp from the leuA start codon. Cycling conditions for both PCRs consisted of an initial denaturing step at 94 °C for 2 min, followed by 30 cycles of 94 °C for 30 sec, 55 °C for 1 min with a 1 sec/cycle decrease in time, and 72 °C for 2 min; with a final extension step of 72 °C for 15 min. The nested PCR fragments were cleaned with sodium acetate/ethanol precipitation and cloned into pGEM-T Easy Vector (Promega, USA). Inserts were confirmed with colony PCR and the clones for each biotype with the longest fragments were sequenced as prescribed by the manufacturer using ABI BigDye v3.1. System (Applied Biosystems, USA) on an ABI 3100 automated sequencer (Applied Biosystems). The longest sequenced fragment that
correctly aligned with the leucine plasmid was regarded as the leader sequence and submitted to Genbank (GU145279 and GU145280).

**The inverted repeat region in the Aphididae**

The large inverted repeat region on the *Buchnera* leucine plasmid was isolated for the different members of the Aphididae using degenerate primers (Aphididae_repA2_For, 5\' - GAAT TAA CDAAAATWGGY C K MARGG-3\' and Aphididae_leuA_Rev, 5\' - CCATTACTAG-TATCCTAATGCTTGRTCNCCATCNCG-3\') designed from multiple sequence alignments of the region between the repA2 and leuA using ClustalW (Larkin et al. 2007). Total DNA (30 ng), with 0.5 U ExSel High Fidelity DNA polymerase (Southern Cross Biotechnology, RSA), 1 × reaction buffer with MgSO	extsubscript{4} (Southern Cross Biotechnology), 100 µM of each dNTP and 0.4 µM of each primer, was used in 25 µL reaction volumes to amplify the inverted repeat region using a GeneAmp 9700 thermocycler (Applied Biosystems). Cycling conditions consisted out of an initial denaturing step at 94 °C for 2 min, followed by 12 touchdown cycles (15 sec at 94 °C, 30 sec at 60 °C, decreasing 1 °C /cycle, 30 sec at 72 °C), 25 standard cycles (10 sec at 94 °C, 10 sec at 50 °C, 30 sec at 72 °C) and a final 7 min step at 72 °C. The PCR products were cleaned through ethanol precipitation and cloned (pGEM-T Easy Vector, Promega) before unidirectional sequencing (ABI BigDye v3.1. System, Applied Biosystems). Sequence assemblies of the original chromatograms were produced using the default settings of ContigExpress (Vector NTI Advance 9, Invitrogen) (Lu & Moriyama 2004) and identities confirmed with a BLAST (Altschul et al. 1990; 1997) analysis against the non-redundant Genbank database (NCBI, http://www.ncbi.nlm.nih.gov/). Only unique sequences obtained for each species were submitted to Genbank (GU145281-GU145289).

**Software analysis**

Secondary structural analysis of the sequences were performed using the Quickfold on the DINAMelt server at 25 °C (Markham & Zuker 2005) and the Kinefold server (Xayaphoummine et al. 2005). Bacterial promoters with their sigma factor binding sites were determined using BPROM (http://softberry.com). Promoter candidates were investigated using neural network promoter predictions (http://www.fruitfly.org/seq_tools/promoter.html) (Reese & Eeckman 1995; Reese 2001) and Hidden Markov Models (http://bioinformatics.biol.rug.nl/websoftware/ ppp/ppp_start.php). The Suite for Computational identification Of Promoter Elements (SCOPE, http://genie.dartmouth.edu/scope/) (Carlson et al. 2007; Chakravarty et al. 2007) and the pattern
discovery tools of the RSA webpage (http://www.bi.up.ac.za/rsa-tools/) (van Helden et al. 2000), were used to identify possible cis and other regulatory elements. WebLogo 3 (http://weblogo.threeplusone.com/) (Crooks et al. 2004) was used to convert consensus sequences to sequence logos. In all the analysis, B. aphidicola, if present in the option list, was always selected as part of an analysis, otherwise E. coli or prokaryote was used.

RT-qPCR

The iScript One-Step RT-PCR kit with SYBR Green (Qiagen) was used for the RT-qPCR reactions, with a final volume of 25 µL, 50 ng total aphid RNA, and PCR conditions in accordance to the manufacturer’s recommendations. Primer and PCR optimizations were done in accordance to the manufacturers protocols (Bio-Rad, USA). All reactions were performed on an iCycler with an iQ Real-Time PCR Detection System (Bio-Rad). Single amplicons were confirmed with melting curve analysis and agarose gels. Primers for the RT-qPCRs were designed using Primer 3 (Rozen & Skaletsky 2000). Previously, Moran et al. (2005) used rpsL (ribosomal protein) as a RT-qPCR standard since it showed constant transcription. Based on this we’ve selected similar genes/subunits available on Genbank for B. aphidicola of the RWA, i.e. RNA polymerase β-subunit (rpoB, AF465521) (rpoB_F, 5’-TACAACGCACGCATTATTC-3’ and rpoB_R, 5’-ACGGTGACTGGAAGTTTTCG-3’) and 16S ribosomal RNA (16S rRNA, M63251) (16SBuDn_F, 5’-TGTAGCGGTGAAATGCGTAG-3’ and 16SBuDn_R, 5’-CC-TCCAAGTCGACATCGTTT-3’).

Previous results on the leucine plasmid (pLeu-Dn) indicated that the only difference between the RWA biotypes investigated here, resided within an inverted repeat region upstream of leuA (Swanevelder et al. 2010). Primers were therefore designed for leuA based on the Genebank accession FJ705299 (leuA_Buch_Dn_F, 5’-TGCATTTTCACATTCTTCTGG-3’ and leuA_Buch_Dn_R, 5’-CTGCAGCTCCTCTGTAGCTCTG-3’). The same study showed that leuA and leuB was in the same open reading frame (ORF), though leuB has its’ own TATAAT box and predicted promoter (Swanevelder et al. 2010). Testing this hypothesis and possible secondary structures due to the leader sequence, we’ve included leuB (leuB_Buch_Dn_F, 5’-TGAATGTGCCATGATTACAGG-3’ and leuB_Buch_Dn_R, 5’-CCCTGAATATCAGGAGCTGAAC-3’) in the analysis.
Results

The leader sequence

Several cloned RACE fragments for the *Buchnera* plasmids with and without the CCC-insert were sequenced and used to produce multiple alignments. The longest sequence in each case was submitted to Genbank (GU145279 and GU145280). The sequenced RACE fragments indicated that the mRNA 5’ untranslated transcription region (5’ UTR) starts on the second stem region of the inverted repeat, just upstream of the *leuA* gene and not as previously suggested to be located before the stemloop structure (Swanevelder *et al.* 2010). It was also found that the leader sequences differ in length (*Figure 4.1*), with longer sequence for the *B. aphidicola* plasmid with the CCC-insert. This finding implies a change to the previously predicted transcription start site. In the new prediction the transcription start site moves upstream, making it part of the conserved stem towards the loop and nearer to the CCC-insert (*Figure 4.1*). This further implies that the predicted *rpoH* (σ^{32}) binding site may be involved in the regulation (Swanevelder *et al.* 2010). However, testing the different predicted promoters (BPROM, Softberry) in a reporter plasmid (pGlow-TOPO Reporter Kit, Invitrogen) within *E. coli* rendered no expression data (data not shown), leading to the conclusion that the high AT-rich *B. aphidicola* predicted promoters tested were either unrecognisable in *E. coli* or alternatively there were too many “recognisable” *E. coli* promoters that interfered with the expression.

The differences in the lengths of the leader sequences produced another possible reason for the observed variance in the transcript expression levels. Secondary structures are known to stabilise the mRNA transcripts of bacteria, thereby increasing their half-life (Emroy *et al.* 1992). We therefore included the predictions of both leader sequences as part of the analysis. The leader sequences produced different 5’ secondary structures (Quikfold, DINAMelt Server) (*Figure 4.1*), with the longer 5’UTR producing structures (*Figure 4.1*, insert B) that have more than double the free energy values than those from the short fragment (*Figure 4.1*, insert A).
Figure 4.1 Buchnera aphidicola of D. noxia with the leader regions (5' UTR) indicated by a blue line for accessions with an upstream CCC-insert and a red line for those without the CCC-insert. Predicted promoters, sigma factor binding sites, AU-rich regions and ribosomal binding sites (RBS), including possible start sites, are indicated. Predicted secondary structures (Quickfold, DINA melt Server, 25 °C) for the leader sequences are given in the inserts A and B. Six structures were predicted for the short 5' UTR (insert A) and three for the longer leader region of the CCC-insert containing Buchnera (insert B). The longer leader sequence produced predicted structures that are approximately -2.3 kcal mol\(^{-1}\) more stable than those of the shorter 5' UTR. These structures could increase the half life of the RNA molecules by preventing RNA degradation.
The inverted repeat in the Aphididae

The preceding results indicate that the second stem of the inverted repeat region, just upstream of the leuA gene, could be part of a functional area within the RWA biotypes investigated. We know from previous work that the inverted region in which the insert is located, is conserved within the family Aphididae (Silva et al. 1998). This suggests that the inverted repeat region, especially the second stem, plays a regulatory role or is preserving a regulatory region. We have attempted to gather support for this hypothesis by sequencing the inverted repeat region of a number of aphid individuals, from one or more localities, in order to obtain data that can indicate the extent to which this region is utilized within species/genera as a regulatory mechanism and possibly help indicating the underlining mechanism.

The region, though conserved in most samples, did show some variation within Buchnera from the same host species (Figure 4.2, in bold red), i.e. for Buchnera from aphids originating from the same population/sampling site, hosts B. brassicae (GU145288-9) and H. pruni (GU145286-7) and from different regions, hosts D. noxia (AF041837, FJ705299 and FJ705301), M. rosae (GU145283 and AJ006881) and U. sonchi (GU145285 and AJ006873). Changes observed in the conserved stem regions usually produce imperfect hairpins, e.g. B. aphidicola from B. brassicae, D. noxia and U. sonchi, resulting in an increase in the Gibbs free energy (ΔG) of app. 5 kcal mol\(^{-1}\) and less stable structures. The only exception was two point mutations in the stem regions of B. aphidicola of H. pruni that had little effect on the predicted structure or stability of the hairpins (GU145286 vs. GU145287, ΔG difference of 0.57 kcal mol\(^{-1}\)). These two point mutations are predicted to interact as T-G base pairings in the stem region (nearest-neighbor effect), thereby preventing major structural changes. Within-hosts differences were also obtained for the variable loop regions of B. aphidicola of D. noxia and M. rosae. A single nucleotide variation in the loop region of the endosymbiont of D. noxia (AF041837, FJ705301) alters the predicted stability of the hairpin by -0.57 kcal mol\(^{-1}\), while the stem and loop region lengths are maintained. Changes in loop length and composition, together with varying stem lengths, can have major implications on the predicted structural stability – though a perfect hairpin is maintained. This is illustrated in the two B. aphidicola accessions of M. rosae where the hairpins’ ΔGs differ by 4.23 kcal mol\(^{-1}\), as a result of the changes.

A comparison of this region within a Buchnera from the same aphid genus (i.e. Diuraphis and Rhopalosiphum) showed that the variations between the species seemed to be mainly focused to variable regions (Figure 4.2: bold blue text), i.e. the loop region and the variable region between the stemloop and the leuA start site, and to the edges of the stem regions, i.e. a core region is conserved within the stems. Indeed, the conserved cores located in the stem region of the hairpin are preserved across all three tribes of the Aphididae (Figure 4.2) just as...
Figure 4.2 The region between the repA2 and leuA on the leucine plasmid of B. aphidicola accessions originating from various aphid hosts. Variations within a species is indicated by bold red lettering, variations within a genus by bold blue lettering and the start site of the 5’ UTR leader sequence by bold gold lettering. The genes repA2 and leuA are shaded in grey with their end and start codons, respectively, in bold. The inverted repeat regions of the family Aphididae are underlined on the consensus sequence and the regions indicated by black arrows. The conserved core region (red double arrows) of the inverted repeat region (black) includes predicted promoters, transcription factor binding sites and start codons (see legend). The Gibbs’ free energy (ΔG) values for the stemloop structures were calculated with Quickfold (DINAMelt Server, at 25°C in kcal mol⁻¹). Predicted stemloop structures (Loop) is given as perfect (P) or imperfect (IP), with the number of additional loops indicated. The number of suboptimal T-G pairs (TG) predicted is also listed. RSA-tools’ Consensus and Convert-matrix programs were used to obtain the core regions of the stemloop structures that are conserved within the family.
previously shown (Silva et al. 1998). This suggests that changes to these variations within the variable loop region and the edges of the inverted repeats could destabilize the stemloop structure in a similar way to the CCC-insert and could theoretically regulate gene expression.

**RT-qPCR**

Relative gene expression levels of two genes on the *pleuABCD* plasmid of *B. aphidicola*, *leuA* and *leuB*, were quantified to assess whether the presence of the CCC-insert had any effect on the expression of these genes (*Figure 4.3*). A difference found in the expression levels has the potential to contribute to RWA adaptation to new hosts via the endosymbiont. It is known that *B. aphidicola* of the South African RWA biotypes have lower plasmid copy numbers (0.35 copies/bacterial chromosome) than their US counterparts (1.04 and 0.88 copies/bacterial chromosome for RWA-US1 and RWA-US2 respectively) (Chapter 3: *Figure 3.4*, Swanevelder et al. 2010). Significantly higher expression levels were obtained for *leuA* and *leuB* after normalization with *rpoB* (*Figure 4.3*) and *16S rRNA* (not shown) in the South African *B. aphidicola* accessions with the CCC-insert than *B. aphidicola* without the insert (*i.e.* RWA-US1 and RWA-US2).
Figure 4.3 The relative gene expression levels per plasmid copy of \textit{leuA} and \textit{leuB} after normalization with \textit{rpoB}. 
Discussion

The role of the endosymbiont in an aphid’s adaptability to new environments has been rarely investigated, though the two organisms are totally inter-dependent on each other for their survival (Munson et al. 1991; Lai et al. 1994; Douglas 1998; Sandstrom & Moran 1999; Baumann et al. 2006). Recently Dunbar et al. (2007) clearly illustrated the influence that the endosymbiont may have on its host by showing that a single point mutation in the bacterium’s genome determined the aphid host’s tolerance to environmental temperature changes. In our previous study, we identified a single mutation (CCC-insert) that differs between RWA biotypes which was located in a proposed regulatory region situated on the leucine plasmid upstream from the leuA-leuB ORF (Silva et al. 1998; Chapter 3; Swanevelder et al. 2010). RWA biotypes and their endosymbionts are known to have little sequence variation (Lapitan et al. 2007; Shufran et al. 2007); thereby suggesting that this CCC-insert may play a functional role in the development of at least some RWA biotypes.

Here we showed that this insert does result in higher transcript copy numbers (Figure 4.3). From our predicted models for the sequenced 5’UTR of leuA (Chapter 3), it was shown that the transcriptional start of the region followed the loop in the second inverted repeat upstream from the leuA gene on the leuABCD plasmid. The sequenced fragments also illustrated that the CCC-insert caused and increased 5’UTR length. This suggests a change in the transcription start sites as a result of the insert, forming a new predicted rpoH (σ32) binding site. Only two transcription factors in B. aphidicola are known, i.e. rpoH that encodes the sigma factor σ32 and rpoD that encodes σ70 (Shigenobu et al. 2000).

The rpoH (σ32) has a functional significance in E. coli as a heat shock transcription factor (TF) (Erickson et al. 1987). In order to understand how a heat shock/stress response TF could regulate gene expression under normal conditions, we investigated the Buchnera genome and the implications it may have on protein function/stability. Buchnera genomes are drastically reduced, have little or no recombination; occur as multiple chromosomal copies per cell; are AT-rich with no codon bias; accrue detrimental mutations and have elevated evolutionary rates with increased nucleotide substitution rates (Moran 1996; Clark et al. 1999; Komaki & Ishikawa 1999; Itoh et al. 2002; Moya et al. 2002). These genomic conditions have resulted in amino acids sequences that produce structurally destabilized or miss-shaped proteins (Moran 1996; Van Ham et al. 2003; Wilcox et al. 2003). The destabilized/deformed proteins initiate the release of the σ32 factor that binds and redirect RNA polymerase to σ32 promoters, thereby regulating the expression of genes that
encodes chaperones, proteins associated with cellular homeostasis restoration and proteases (Wilcox et al. 2003). It is suggested that the endosymbiont attempts to retain protein functionality by stabilising the protein structures using “heat shock” chaperonins. However, unlike a temporarily protein destabilizing environment, it seems that the genome of Buchnera continuously produce miss-formed proteins, thereby necessitating a steady level of protein chaperonin production under normal conditions. Indeed, the chaperonin GroEL constitutes about 10 % of the cellular proteins under normal growth conditions in the Buchnera of Schizaphis graminum (Baumann et al. 1996). The role of this chaperonin in protein stabilization is not only underlined by the positive selection pressure under which the protein is maintained (Fares et al. 2002), but also by experimental evidence that showed that with its (GroEL) co-expression Buchnera enzyme function is enhanced (Huang et al. 2008). Buchnera, even though it has lost the ability to regulate most of the heat stress proteins under heat stress conditions (Baumann et al. 1996; Sato & Ishikawa 1997a; b; Wilcox et al. 2003), still retains a σ32 TF that has the necessary binding domains needed for functionality (Wilcox et al. 2003). A low, but continuous expression of rpoH under normal growth conditions is therefore necessary to maintain the steady production of the various required protein chaperonins to ensure protein stability. Indeed, rpoH is only slightly up-regulated from its norm under heat shock conditions (Wilcox et al. 2003).

In order to understand how this new putatively formed TF could regulate gene expression, we investigated the expression of leuA and leuB in RWA biotypes, with and without the CCC-insert. We found a twofold difference in the expression of leuA and leuB as expressed per copy number (Figure 4.3).

The newly formed rpoH (σ32) binding site may explain the higher levels of transcripts observed in the biotypes with the introduced TFBs. However, the length differences in the leader sequences produced another possibility for the higher transcript levels. Secondary structures are known to play a role in stabilising mRNA transcripts in bacteria (Emroy et al. 1992). Here the predicted 5’ secondary structures of the longer 5’UTR (Figure 4.1 insert B) are all more than twice as stable as those produced by the shorter leader sequence (Figure 4.1 insert A). This suggests that secondary structures in the leader sequence may provide more stability to Buchnera transcripts originating from plasmids with the CCC-insert, thereby increasing its half life, and thus the total transcript levels. Secondary structures in the translation initiation region is also known to play a role in translation regulation, even within B. aphidicola (Tchufistova et al. 2003). However, the conserved translational control that features in the leader sequence necessary for S1 protein
translational control (Tchufistova et al. 2003) is absent from the leader sequences identified here (Figure 4.1).

The functional mutation in the inverted repeat region of *B. aphidicola*, together with the conservation of the structure within the Aphididae (Silva et al. 1998), supports a wider functional role within the family. The observed changes never occurred within the conserved “core” region of the inverted repeats, but were kept to variable regions within the stemloop structure. This low level of variation observed could be due to the small species sample size used here. However, changes observed usually did alter the structures of predicted hairpins, thereby increasing/decreasing its stability, e.g. the *B. aphidicola* of *M. rosae* had hairpin $\Delta G$s that differ with 4.23 kcal mol$^{-1}$ even though a perfect stemloop was maintained. Within a genus, changes were also never within the conserved core. If the 5’ UTRs obtained for the *B. aphidicola* of *D. noxia* biotypes are the norm for the family, the highly conserved core of the second inverted repeat is a likely promoter region. This would suggest that the stemloop structure is used to protect the promoter region, i.e. the core regions, while mutations that affect its structure and stability is used to control gene expression or the half-life of the transcript. Therefore, it can be argued that *Buchnera* utilizes structural changes in this region to control gene expression via structural stability, i.e. easier access to start sites, or via structural changes in leader sequences that may increase transcript half-lives.

## Conclusion

The initial transfer of the leucine biosynthetic genes, from the chromosome to a plasmid, was probably one of the main reasons for a successful aphid-*Buchnera* symbiosis (Latorre et al. 2005). However, different aphids require different levels of this essential amino acid. Indeed, the RWA has the ability to increase certain essential amino acids, including leucine, in susceptible hosts’ phloem, but cannot achieve the same in a resistant host (Telang et al. 1999; Porter & Webster 2000; Sandstrom et al. 2000; Ni et al. 2001). Previously it was believed that the fine tuning of leucine production in an aphid species was done through changes in the plasmid copy number (Thao et al. 1998; Plague et al. 2003; Latorre et al. 2005). However, the increase in the *leuA-leuB* ORF transcripts relative to the known plasmid copy numbers, suggest that this could be a form of regulation within the species. The existence of variable regions within an aphid species and differences in structural stabilities of either the plasmid or leader sequences within *Buchnera* plasmid could arguably support a regulatory mechanism for leucine control. The fact that the same
insertion occurred twice, independently and involved multiple nucleotides that are not part of the
*Buchnera*’s genome bias (AT-rich genome) (Chapter 3; Swanevelder et al. 2010), further supports
this hypothesis. We therefore propose that the variation within the inverted repeat region, together
with plasmid copy numbers, are used by *Buchnera* to control gene expression, either through higher
expression levels or via 5’ UTR mRNA stabilization.

We showed that copy number is not necessary the same as expression level in *Buchnera*. This
suggests that the *Buchnera* endosymbiont is employed by the RWA to compensate for lower leucine
levels when it is feeding on resistant cultivars. Since RWA biotypes are characterized based on their
ability to overcome different hosts’ resistances rather than aphid anatomy and morphology, an
endosymbiont mutation that allows feeding on previously resistant cultivars, could result in a
classifiable “new RWA biotype”.

**Acknowledgements**

We would like to thank Prof N.L.V. Lapitan (Colorado State University) for the USA RWA
biotypes, Dr. V. Tolmay (Small Grains Institute) for the SA RWA biotype, Dr. G.J. Puterka (USDA-
ARS, Stillwater, Oklahoma, USA) for the other *Diuraphis* species and Mr. I. Millar (South African
National Collection of Insects (SANC), PPRI) for the other members of the Aphididae. This
research was funded by THRIP (Technology and Human Resources for Industry Programme, South
Africa) and the Winter Cereal Trust.

**References**

Altschul, S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman (1990) Basic local alignment

Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller and D. J. Lipman

characterization of plasmids containing genes encoding enzymes of leucine biosynthesis in


Sato, S. and H. Ishikawa (1997a) Expression and control of an operon from an intracellular symbiont which is homologous to the *groE* operon. *Journal of Bacteriology* 179: 2300-2304.


Thao, M. L., L. Baumann, P. Baumann and N. A. Moran (1998) Endosymbionts (*Buchnera*) from the aphids *Schizaphis graminum* and *Diuraphis noxia* have different copy numbers of the plasmid containing the leucine biosynthetic genes. *Current Microbiology* **36**: 238-240.


