Conservation of the binding site for the arginine repressor in all bacterial lineages

ABSTRACT

We found that the ArgR/AhrC recognition signal is conserved in all genomes that contain genes encoding orthologous transcription factors of this family. All genomes studied except M. tuberculosis contain ABC transport cassettes (related to the Art system of E. coli) belonging to the candidate arginine regulons.

INTRODUCTION

Background Bacterial and archaeal transcriptional regulators typically form large protein families consisting of numerous paralogs (for example the LacI/GntR, AraC and DeoR families). Only three readily detectable clusters of orthologous transcription factors include just one or two representatives from a broad range of diverse branches of bacteria, namely the SOS repressors LexA/DinR, the heat-shock repressor HrcA, and the arginine repressor ArgR/AhrC (Table 1). A comparison of the coevolution of these conserved regulators and their binding sites in DNA could reveal general trends in the evolution of regulons. The signals recognized by LexA in Gram-negative bacteria and by its ortholog DinR in Gram-positive bacteria (the SOS box and the Cheo box, respectively) are completely different. Accordingly, the DNA-binding domains of these proteins are divergent (Table 1). The heat-shock regulator HrcA binds CIRCE elements that are located upstream of genes encoding heat-shock proteins (molecular chaperones) in many different genomes; in the mycoplasmas, HrcA also regulates heat-shock protease genes. The CIRCE signal is very specific (two complementary nonamers with a 9 base pair (bp) spacer) and is extremely highly conserved in all genomes that encode HrcA (not more than five, and usually less than three, mismatches to the consensus in all known and predicted sites). The amino acid sequence of HrcA is conserved as well (Table 1). The arginine regulon, which is regulated by the arginine repressor ArgR/AhrC, represents an evolutionary strategy distinct from that of either the SOS or the heat-shock regulons. The DNA-binding domains of the ArgR/AhrC family are less conserved than those of the HrcA family, but more conserved than those of the LexA/DinR family (Table 1, column 5). DNA signals recognized by ArgR/AhrC are also similar in several bacterial lineages at least. These sites often occur in pairs, although single-box sites have also been shown to bind ArgR/AhrC, for example the sites in the catabolic operons of B. subtilis, the adenine deaminase pathway operon in Bacillus licheniformis, and the cer recombination region of the E. coli plasmid ColE1 (; see also the study of mutated ArgR). Unlike the CIRCE element, the ARG box seems to be weakly conserved, even within a genome, and the specificity of recognition is often achieved by cooperative interactions between tandem sites, as shown in both experimental and statistical studies. The set of ARG boxes from different genomes, however, is fairly homogeneous, and indeed, arginine repressors from different bacteria appear to be at least partially interchangeable within major taxonomic groups: there is some cross-binding between ArgR and AhrC; ArgR but not AhrC binds to the Thermus thermophilus sites and AhrC binds to the Streptomyces coelicolor sites. The ARG box consensus was described as TNTGAATWWWWATTCANW in E. coli, CATGAATAAAAATKCAAK in B. subtilis and AWTGCATRWWYATGCAWT in Streptomycetes (where W = A or T, K = G or T, R = A or G, Y = T or C, N = any base; Table 1). In addition, binding of ArgR homologs to the sites similar to ARG boxes was

reported for other Bacillus species (B. licheniformis and B. stearothermophilus), and for Salmonella typhimurium. Several ArgR-binding sites were predicted on the basis of similarity with the E. coli consensus in the upstream regions of various genes involved in arginine metabolism in Moritella. In a previous study, we used comparative genomic analysis of regulatory signals to predict the gene composition of the arginine regulon of Haemophilus influenzae using the well characterized E. coli regulon as the starting point. Here we extend this analysis to explore the conservation of the ARG box in all bacteria that encode an ortholog of the ArgR repressor.

CONCLUSION

Conclusions The composition of the ARG regulons in different bacteria is known to vary mainly because of diversity in the arginine degradation pathways and species-specific paralogs. The question of the origin of 'additional' ARG boxes thus arises. Because of the low conservation of the ArgR-binding signal, it is possible that some of the sites could be convergent in origin. Moreover, each genome contains a large number of potential ARG box-like sequences that could become actual sites when they become located upstream of an arginine metabolism gene following chromosomal rearrangements. In contrast, CIRCE elements appear to be direct descendants of the ancient regulon present in the common ancestor of the Bacteria, because the variation in the composition of the CIRCE regulon is minimal and the few additional sites found in some genomes are apparently products of duplication. Most other DNA-binding domains of transcriptional regulators (including LexA) seem to undergo considerable changes together with their DNA signals and regulons. Thus, the evolution of the arginine regulon and ARG boxes seems to reflect a tradeoff between maintaining regulon flexibility on one hand and retaining the universal regulatory mechanism on the other. Another interesting aspect of the arginine regulon strategy is the use of single and cooperative sites. In E. coli, the use of cooperative binding sites by ArgR seems to be a consequence of a requirement for a sharper response to a stimulus (arginine starvation) compared to the SOS response (single sites are usually used by LexA). Unfortunately, the available data seems to be insufficient to draw any systematic conclusions. In particular, as second sites in the cooperative cassettes are often weak (have low scores), some of them could be missed by the recognition rule. Direct experimental studies are needed to clarify this issue. Another problem that was not directly addressed in this study is the role of the E. coli arginine repressor in recombination and its binding to the cer site, which contains a single ARG box. We have noted, however, conservation of this box in the monomerization site ckr of the plasmid ColK. There are a few more transcription factor families (biotin operon repressor, COG1654; putative stress-responsive transcriptional regulator PcpC, COG1983; Bvg accessory factor homologs, COG1521) with a single representative per genome, and it would be interesting to compare them as well. They do not, however, contain a sufficient number of experimentally determined binding sites and are not so ubiquitous in the bacterial genomes as the three regulators discussed previously. With more available genomes, we hope that our approach, combined with positional analysis aimed at finding co-localized, and thus possibly functionally related enzymes and regulator genes, will enable us to make this comparison. On the other hand, we feel that the predictions made in this study, especially identification of the Art family ABC transporters in several diverse genomes, are sufficiently interesting to warrant experimental verification.