

RESEARCH ARTICLE

Open Access

Comparative analysis of diguanylate cyclase and phosphodiesterase genes in *Klebsiella pneumoniae*

Diana P Cruz^{1,2}, Mónica G Huertas^{1,3}, Marcela Lozano¹, Lina Zárate¹ and María Mercedes Zambrano^{1*}

Abstract

Background: *Klebsiella pneumoniae* can be found in environmental habitats as well as in hospital settings where it is commonly associated with nosocomial infections. One of the factors that contribute to virulence is its capacity to form biofilms on diverse biotic and abiotic surfaces. The second messenger Bis-(3'-5')-cyclic dimeric GMP (c-di-GMP) is a ubiquitous signal in bacteria that controls biofilm formation as well as several other cellular processes. The cellular levels of this messenger are controlled by c-di-GMP synthesis and degradation catalyzed by diguanylate cyclase (DGC) and phosphodiesterase (PDE) enzymes, respectively. Many bacteria contain multiple copies of these proteins with diverse organizational structure that highlight the complex regulatory mechanisms of this signaling network. This work was undertaken to identify DGCs and PDEs and analyze the domain structure of these proteins in *K. pneumoniae*.

Results: A search for conserved GGDEF and EAL domains in three sequenced *K. pneumoniae* genomes showed that there were multiple copies of GGDEF and EAL containing proteins. Both single domain and hybrid GGDEF proteins were identified: 21 in *K. pneumoniae* Kp342, 18 in *K. pneumoniae* MGH 78578 and 17 in *K. pneumoniae* NTUH-K2044. The majority had only the GGDEF domain, most with the GGEEF motif, and hybrid proteins containing both GGDEF and EAL domains were also found. The I site for allosteric control was identified only in single GGDEF domain proteins and not in hybrid proteins. EAL-only proteins, containing either intact or degenerate domains, were also identified: 15 in Kp342, 15 in MGH 78578 and 10 in NTUH-K2044. Several input sensory domains and transmembrane segments were identified, which together indicate complex regulatory circuits that in many cases can be membrane associated.

Conclusions: The comparative analysis of proteins containing GGDEF/EAL domains in *K. pneumoniae* showed that most copies were shared among the three strains and that some were unique to a particular strain. The multiplicity of these proteins and the diversity of structural characteristics suggest that the c-di-GMP network in this enteric bacterium is highly complex and reflects the importance of having diverse mechanisms to control cellular processes in environments as diverse as soils or plants and clinical settings.

Keywords: *Klebsiella pneumoniae*, Biofilm, Diguanylate cyclase, Phosphodiesterase, c-di-GMP

Background

Klebsiella pneumoniae, an opportunistic pathogen responsible for a wide range of nosocomial infections that include pneumonia, bacteremia and urinary tract infections, is estimated to cause approximately 8% of hospital acquired infections [1-5]. This Gram-negative bacterium

can also be found in the environment in association with plants, as well as in soil and in water [2,6]. One important factor associated with virulence in *K. pneumoniae* is its capacity to adhere to surfaces and form biofilms. Although the formation of biofilms by *K. pneumoniae* is still not fully understood, several key determinants have been identified such as pili, polysaccharides, quorum sensing and transport and regulatory proteins [7-13]. More recently, it has been shown that c-di-GMP controls type 3 fimbria expression and biofilm formation in

* Correspondence: mzambrano@corpogen.org

¹Molecular Genetics, Corporación Corpogen, Carrera 5 No. 66A-34, Bogotá, Colombia

Full list of author information is available at the end of the article

K. pneumoniae by binding to and modulating the activity of the transcriptional regulator MrkH [14,15]. The second messenger c-di-GMP is known to play a key role in several cellular functions as well as in biofilm formation in bacteria where it modulates the transition between planktonic and sessile lifestyles. Low levels of c-di-GMP result in increased motility while high levels promote adhesion to surfaces, production of exopolysaccharides and biofilm formation [16,17].

The intracellular levels of c-di-GMP are regulated by the antagonistic activity of diguanylate cyclase (DGC) enzymes and phosphodiesterases (PDEs) that catalyze synthesis and hydrolysis of this molecule, respectively [16,18]. Several genetic and biochemical studies have shown that besides their C-terminal catalytically active A site, most of these proteins harbor N-terminal sensory domains that can respond to different internal and external signals, triggering activation of DGCs or PDEs. When enough c-di-GMP is available, it binds different effector molecules, proteins or RNAs, which influence cell behavior [18]. The active site of DGCs contains a conserved GGDEF domain, characterized by the GG(D/E)EF motif, while PDE activity is associated with C-terminal EAL or HD-GYP domains [16,17]. These domains can be found separately or together, forming hybrid proteins that have both GGDEF and EAL domains. Hybrid proteins usually have either PDE or DGC activity, although in some cases both functions are apparently present [17,18]. DGCs can also be subject to allosteric product inhibition by c-di-GMP, which binds to a secondary site (I site) separated from the A site by 5 amino acids [16]. This feedback control helps to maintain adequate pools of c-di-GMP, avoiding excessive consumption of the GTP substrate and reducing stochastic perturbations in cellular c-di-GMP content [16,17]. GGDEF and EAL proteins can also contain one or more transmembrane regions and signal peptides that can anchor these proteins to the membrane, most probably allowing physical isolation of different GGDEF and EAL systems to unique microenvironments [17]. In addition, some bacterial species can harbor multiple copies of proteins with GGDEF and EAL domains. Many of these copies may contain degenerate sites that are inactive and do not directly synthesize or degrade c-di-GMP but have adopted alternative functions, either as c-di-GMP binding effector proteins or through direct macromolecular interactions with no involvement of c-di-GMP at all [17]. The diversity of sensor domains coupled to the multiplicity of these genes reveal a complex c-di-GMP network that integrates diverse environmental and cellular signals [16,17].

This work was carried out to identify GGDEF and EAL domain-containing genes in three sequenced *K.*

pneumoniae genomes. Searches were done for the conserved GGDEF/EAL domains and the RxxD allosteric I site. Sensory domains associated with these proteins, as well as transmembrane helices and signal peptides were also identified. The results show that there are multiple copies of these genes in the sequenced genomes studied and that some of these are shared while others are unique to a particular strain.

Results and discussion

Multiplicity of genes encoding GGDEF and EAL containing proteins

To have an inventory of the number of genes coding for GGDEF and EAL domain-containing proteins, PSI-BLAST was used to identify the conserved GG(D/E)EF and E(A/V)L motifs in the three sequenced *K. pneumoniae* genomes. The genomes available at the time this analysis was done included one environmental strain, *K. pneumoniae* Kp342, a nitrogen-fixing endophyte isolated from corn [6], and two clinical isolates from the same subspecies: *K. pneumoniae* subsp. *pneumoniae* MGH 78578, isolated from a patient with nosocomial pneumonia [6], and *K. pneumoniae* subsp. *pneumoniae* NTUH-K2044, isolated from a patient with a hepatic abscess and meningitis [19]. All genomes had multiple copies for proteins with GGDEF domains: 17 for NTUH-K2044, 18 for MGH 78578 and 21 for the environmental isolate Kp342 (Table 1). The majority of these proteins contained the GGDEF sequence motif and only 30% had GGDEF (Figure 1). A subset of the proteins (29%) had both GGDEF and EAL domains and more than 50% of these had GGDEF degenerate domains. Two GGDEF-only proteins (KPK_A0039 and KPN_pKPN3p05901) had GGDEF degenerate domains and were found on plasmids. Multiple copies of proteins with single EAL domains were also identified: 15 for the environmental isolate Kp342, 15 for MGH 78578 and 10 for NTUH-K2044 (Table 1). Most of these proteins (61%) had an intact EAL domain, including the EVL motif (Figure 1), and 39% had EAL degenerate domains (Table 1). Some of the EAL degenerate proteins, such as KPK_A0040 and KPN_pKPN3p05966, were found on plasmids.

To further characterize these proteins, signal peptides, sensor and conserved domains were identified. Only 5% of GGDEF and 7% of EAL proteins in *K. pneumoniae* included signal peptides (Table 1), indicating that they could be transported across or anchored in membranes [20,21]. A larger proportion of the proteins contained transmembrane segments, 73% of the GGDEF and 57% of EAL-containing proteins (Table 1), suggesting that regulation and/or enzyme activity is most likely occurring at the membrane, as has been suggested [22,23].

Table 1 List of domains found in the genomes of *K. pneumoniae* 342, MGH 78578 and NTUH-K2044

| | <i>K. pneumoniae</i> 342 | | | <i>K. pneumoniae</i> MGH 78578 | | | <i>K. pneumoniae</i> NTUH-K2044 | | | Total Predicted DGC | Total Predicted PDE |
|------------------------|--------------------------|-------------|-----|--------------------------------|-------------|-----|---------------------------------|-------------|-----|---------------------|---------------------|
| | GGDEF | GGDEF + EAL | EAL | GGDEF | GGDEF + EAL | EAL | GGDEF | GGDEF + EAL | EAL | | |
| Total proteins | 15 | 6 | 15 | 13 | 5 | 15 | 12 | 5 | 10 | 56 | 56 |
| Transmembrane segments | 11 | 5 | 6 | 9 | 4 | 7 | 8 | 4 | 6 | 41(73%) | 32 (57%) |
| Signal peptides | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 3 (5%) | 4 (7%) |
| With GAF domain | 3 | 0 | 1 | 4 | 0 | 0 | 3 | 0 | 0 | 10 (18%) | 1 (2%) |
| With HAMP domain | 2 | 1 | 0 | 2 | 1 | 0 | 1 | 1 | 0 | 8 (14%) | 3 (5%) |
| With PAS domain | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 6 (11%) | 3 (5%) |
| With BLUF domain | 0 | 0 | 3 | 0 | 0 | 2 | 0 | 0 | 2 | 0 | 7 (12%) |
| With MASE domain | 2 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 5 (9%) | 6 (11%) |
| With CACHE domain | 1 | 0 | 0 | 2 | 0 | 0 | 2 | 0 | 0 | 5 (9%) | 0 |
| With CHASE domain | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 (2%) | 1 (2%) |
| With CSS-motif domain | 0 | 0 | 5 | 0 | 0 | 6 | 0 | 0 | 5 | 0 | 16 (28%) |
| With sensor domains | 9 | 4 | 10 | 9 | 3 | 9 | 7 | 3 | 8 | 35 (62%) | 37 (66%) |
| With allosteric I site | 7 | 0 | 0 | 7 | 0 | 0 | 5 | 0 | 0 | 19 (34%) | 0 |
| With degenerate GGDEF | 1 | 3 | 0 | 1 | 3 | 0 | 0 | 3 | 0 | 11 (20%)* | 9 (56%)* |
| With degenerate EAL | 0 | 2 | 6 | 0 | 2 | 6 | 0 | 2 | 4 | 6 (38%)* | 22 (39%) |

*Average calculated based only on the number of hybrid proteins (16). Numbers in parenthesis indicate % of total for either DGCs or PDEs.

Sensor domains found in GGDEF and EAL containing proteins

One of the most intriguing aspects of the enzymes involved in modulating intracellular levels of c-di-GMP is their modular structure characterized by the presence of additional input sensory domains [24]. Therefore, a search was carried out for the diverse periplasmic, cytoplasmic, and integral membrane domains that have been

described [23,25]. Most of the GGDEF and EAL-containing proteins in *K. pneumoniae* contained sensor domains, 62% and 66%, respectively (Table 1). Some domains were found exclusively or predominantly in GGDEF proteins (CACHE, PAS and GAF) or EAL proteins (BLUF and CSS), while others were shared or found in hybrid proteins (HAMP, CHASE and MASE) [Additional file 1]. As in other bacteria, the different

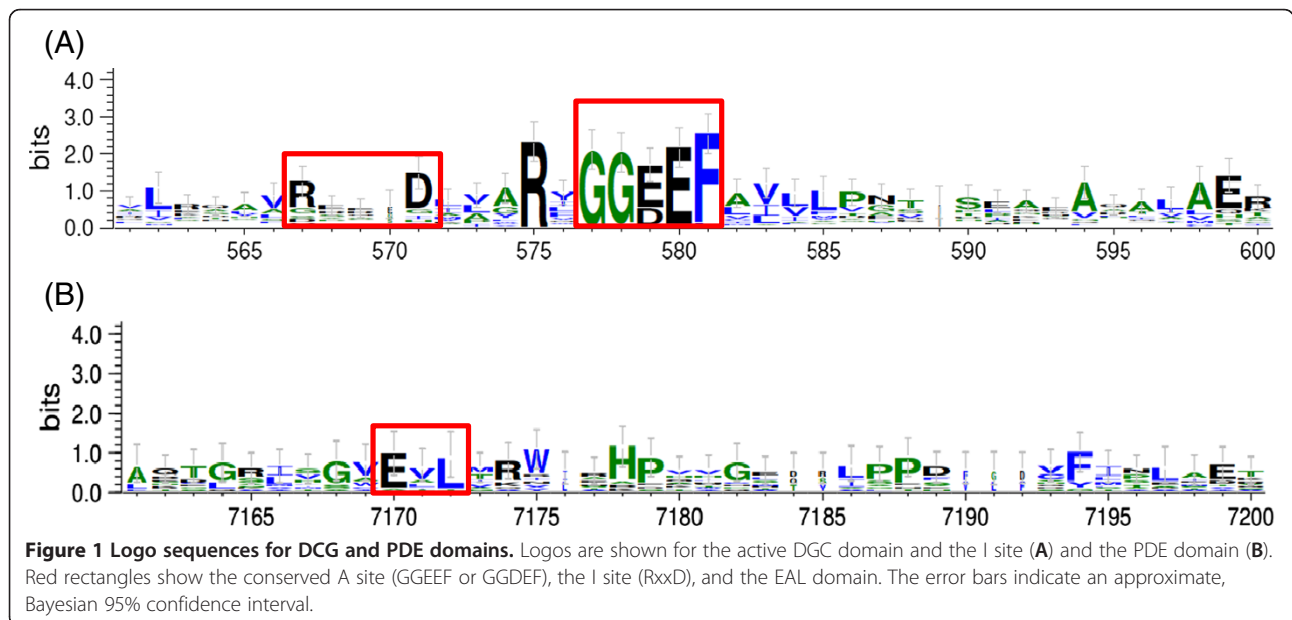


Figure 1 Logo sequences for DGC and PDE domains. Logos are shown for the active DGC domain and the I site (A) and the PDE domain (B). Red rectangles show the conserved A site (GGDEF or GGDEF), the I site (RxxD), and the EAL domain. The error bars indicate an approximate, Bayesian 95% confidence interval.

sensor domains suggest a diverse range of environmental stimuli involved in regulatory responses in this bacterium [26,27] (Table 1). In GGDEF proteins the most frequently found domain was GAF (18%) (cGMP phosphodiesterase, adenylyl cyclase), a cytoplasmic sensor domain that can bind a number of small molecules including monocyclic nucleotides and oxygen and that is also common in signal transducing photoreceptor proteins such as phytochromes, which covalently link chromophores [28]. This was followed by HAMP (Histidine kinases, Adenylyl cyclases, Methyl binding proteins, Phosphatases) domain-containing proteins (14%). This domain has been found in many transmembrane receptors where it transmits signals from periplasmic sensor domains to cytoplasmic output domains via conformational changes [25,29]. The PAS (PER, ARNT and SIM) domain was found only in 11% of the GGDEF proteins. PAS is structurally similar to GAF and can bind small molecules such as heme, flavin, and adenine [29,30]. Other domains were also found in smaller proportions. The membrane-embedded MASE (Membrane-associated sensor) domain [25] was identified in 9% of the GGDEF proteins and 11% of the EAL proteins (Table 1), and the extracellular CHASE (cyclase/histidine kinases-associated sensing extracellular) and CACHE (Ca²⁺ channels and chemotaxis receptors) domains were found in 2% and 9% of the cases, respectively. The CHASE domain apparently recognizes short peptides and cytokines [25,30,31]. The CACHE domain is involved in binding small ligands such as amino acids, sugars and organic acids, and has been found in prokaryotic chemotaxis receptors and animal ion channels [30,31]. The most common sensor domain in EAL proteins was the CSS-motif (28%) of unknown function, followed by BLUF (for 'sensing blue-light using FAD') (12%), which is involved in sensing blue-light and possibly redox states [32]. Some sensor domains identified in other bacteria were not found in *K. pneumoniae*, as was the case for REC (receiving domain with phosphoacceptor site), which is implicated in activation of DGC proteins in organisms such as *Caulobacter crescentus* and *Pseudomonas aeruginosa* [27].

Predicted catalytic activity in GGDEF-containing proteins

Active DGCs consist of two subunits, each with an A site that binds a GTP molecule at the interface between the two subunits. The A site has the characteristic conserved GGDEF or GGEEF motif and point mutations that affect this sequence abolish enzymatic activity [17]. Many DGCs are also subject to allosteric inhibition, which involves binding of c-di-GMP to the I site characterized by the RxxD motif [16,17]. Mutations of the R residue alter the inhibitory function and allosteric control, while mutations of the D amino acid do not [16]. In *K. pneumoniae* 80% of the identified GGDEF-containing proteins

had an intact conserved A site (Figure 1) and of these, only 34% had the conserved I site motif (RxxD) (Figure 1, Table 1), which was present only in single-domain GGDEF proteins. Interestingly, the majority of the proteins that lacked the I site had the GGDEF sequence, which is less common in single-domain DGC proteins. In an analysis of DGC proteins in 867 prokaryotic genomes, about 66% of the DGC single-domain proteins had the GGDEF motif [33]. It has been shown that, in general, I sites are less common in catalytically active DGC hybrid proteins, which has led to the hypothesis that these proteins have lower activities compared to single-domain DGCs, sparing them the need for an I site [33]. Furthermore, 20% of the proteins (11 copies) were found to have degenerate GGDEF domains, two of which, were single-domain GGDEF proteins (KPK_A0039 in Kp342 and KPN_pKPN3p05901 in MGH 78578) [See Additional file 1]. Other hybrid proteins with a degenerate GGDEF domain included KPK_0227 in Kp342, and its homologs in the clinical strains, that had a conserved EAL domain, and proteins KPK_1394 and KPK_0458 in Kp342, and their homologs in the other two strains, that had degenerate GGDEF and EAL domains. Some of these proteins also had additional domains like HAMP and MASE.

Several GGDEF degenerate proteins have been studied in other bacteria. They usually lack DGC activity but in many cases have adopted different functions, some of which involve binding of c-di-GMP [33]. The LapD protein in *Pseudomonas fluorescens*, for instance, has degenerate and enzymatically inactive GGDEF and EAL domains but acts as a c-di-GMP effector protein that modulates biofilm formation. The binding of c-di-GMP to its degenerate EAL domain induces conformational changes of its HAMP domain, resulting in the secretion and localization of the LapA adhesin required for attachment and biofilm formation [34]. Protein CC3396 from *C. crescentus* is a hybrid protein that harbors a degenerate GGDEF domain that is able to bind GTP and subsequently activate PDE activity in the associated EAL domain [35]. Characterization of the degenerate GGDEF proteins in *K. pneumoniae* might therefore reveal interesting novel functions in this bacterium.

Comparative analysis of GGDEF and EAL containing genes

We next compared the GGDEF and EAL-encoding genes in the three sequenced genomes available. There were 15 genes for GGDEF proteins common to all genomes, which had more than 90% identity at the amino acid level (Figure 2). The shared genes could be involved in diverse phenotypes important for cell growth and survival in different environments, some of which could be important for virulence properties, as has been described in other bacterial pathogens [24]. Interestingly, the gene for YfiN (KP1_4180), a protein recently found to have

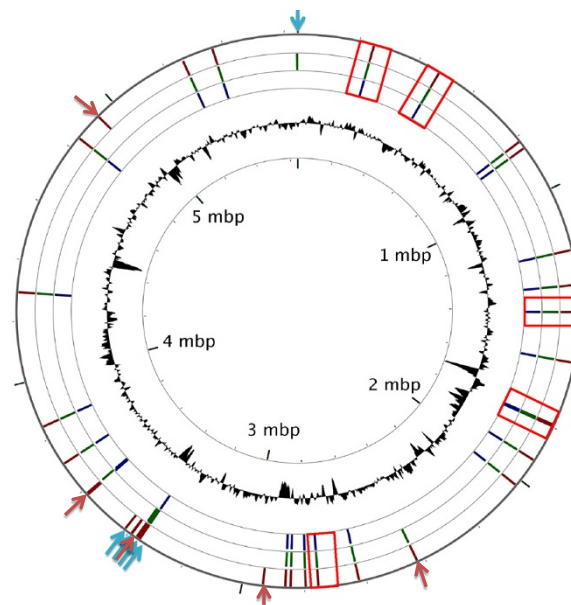


Figure 2 DGCs and PDEs present in the genomes of *K. pneumoniae* 342, MGH 78578 and NTUH K2044. The distribution of GGDEF and EAL domain-containing proteins is shown. The circles represent each genome with lines indicating the DGC and PDE present: red lines for *K. pneumoniae* 342, green lines for MGH 78578 and blue lines for NTUH-K2044. The inner-most circle shows genome positions and the next to last circle shows the GC content. Arrows indicate exclusive copies or copies found in only two of the three genomes, blue arrows for PDEs and red arrows for DGCs, and rectangles represent hybrid proteins with GGDEF and EAL domains. The circular map was generated using the CGView Server [36], with the following parameters: blastx, expect=0.00001, alignment_cutoff=85, identity_cutoff=85.

catalytic activity and to be implicated in pili production and biofilm formation [15], was found in all genomes. Several studies have also shown that environmental *Klebsiella* isolates can be as virulent as clinical strains [2], indicating that they harbor determinants involved in pathogenesis. Four of these GGDEF-containing proteins, one from the environmental strain Kp342 (KPK_A0039), two from strain MGH 78578 (KPN_pKPN3p05967 and KPN_pKPN3p05901) and one from strain NTUH-K2044 (pK2044_00660) were plasmid encoded [See Additional file 1]. Of these, only KPK_A0039 had a homologous gene in the chromosome of Kp342, while KPN_pKPN3p05967, KPN_pKPN3p05901 and pK2044_00660 were unique genes in their respective strains. These genes could therefore have been acquired through horizontal gene transfer, a mechanism common in acquisition of drug resistance in *K. pneumoniae* clinical strains. Of the three, the gene (KPN_pKPN3p05901) had degenerate A and I sites and probably lacks catalytic activity; alternative functions, such as being a c-di-GMP effector protein, would have to be further analyzed.

In addition to shared genes for GGDEF proteins, there were three genes exclusive to the environmental strain Kp342 (KPK_3356, KPK_4891 and KPK_2890) and two additional genes in this strain (KPK_3558 and KPK_3323) that had homologs in only one of the other two genomes analyzed (Figure 2). Gene KPK_3558 had 99% identity at

the amino acid level with gene KP1_1983 of *K. pneumoniae* NTUH-K2044, and KPK_3323 had 98% amino acid identity with gene KPN_01163 from *K. pneumoniae* MGH 78578. The three copies found exclusively in the environmental strain Kp342 could be important for interactions with plants and the capacity to grow as a plant endophyte. In this respect, strain MGH78578 has been reported to have a limited capacity to colonize plant roots in comparison with the environmental strain Kp342 [6]. Thus, the GGDEF containing proteins found in the environmental strain could provide it with additional regulatory and functional versatility.

Although most of the PDE proteins containing the E (A/V)L motif in *K. pneumoniae* were also common to the three genomes, there were unique genes in the environmental strain Kp342 (KPK_3392 and KPK_3355) (Figure 2) and in *K. pneumoniae* MGH 78578 (KPN_00268, KPN_pKPN3p05961, KPN_pKPN4p07065 and KPN_pKPN3p05966), the latter three genes encoded on plasmids (Figure 2) [See Additional file 2]. In Kp342 one gene (KPK_A0040) was found on plasmid pKP187 and had a homolog on the chromosome, and two additional genes (KPK_3327 and KPK_2809) had homologs in only one of the other two genomes. PDE activity in *K. pneumoniae* has been demonstrated only in a few cases: MrkJ (KP1_4554) and BlrP1 (KPN_01598) [13,15]. From our analysis it therefore appears that the environmental strain Kp342 has more

copies of GGDEF/EAL proteins than the clinical isolates. Future studies focused on the function of many of these DGC and PDE genes might shed light on the processes involving growth and survival of this bacterium under different environmental settings.

To further analyze the GGDEF proteins in *K. pneumoniae*, we constructed a phylogenetic tree using protein sequences from *K. pneumoniae* and other bacteria (Figure 3). This analysis showed that most of the GGDEF proteins grouped with proteins from other organisms and not with one another. However, KPK_3356, which is unique in the Kp342 genome, was closely related to KPK_A0039 and had 96% amino acid sequence identity. Interestingly, KPK_A0039 is on plasmid pKP187 of the same strain Kp342 [See Additional file 1] and could therefore have resulted from an event of horizontal gene exchange and a transfer between the plasmid and the chromosome. Other unique GGDEF proteins in Kp342, like KPK_4891 and KPK_2890, were close to GGDEF proteins from *Enterobacter sp.*, with more than 96% amino acid sequence identity (Figure 3). The GGDEF proteins KPN_pKPN3p05967 and KPN_pKPN3p05901,

found on plasmid pKPN3 of MGH78578, also grouped with GGDEF proteins of *Enterobacter sp.*, whereas pK2044_00660, found on plasmid pK2044 of NTUH-K2044, grouped with GGDEF proteins from *Shigella sp.* (Figure 3). These results suggest that many of these proteins are phylogenetically related, perhaps because they are derived from a common ancestor or due to horizontal gene transfer events between *K. pneumoniae* and other bacteria [37]. Additional studies would need to be carried out to further understand the diversity and distribution of GGDEF proteins in these organisms.

Conclusions

As in other enteric bacteria, *K. pneumoniae* harbored multiple copies of GGDEF and EAL-containing proteins. Recent studies have elucidated functions associated with some of these proteins, but much remains to be known in terms of their regulation and involvement in specific cellular functions. Some of the sensor domains identified, such as MASE, CHASE, CACHE and the CSS-motif have not been well characterized to date. In contrast to other well-studied microorganisms, such as

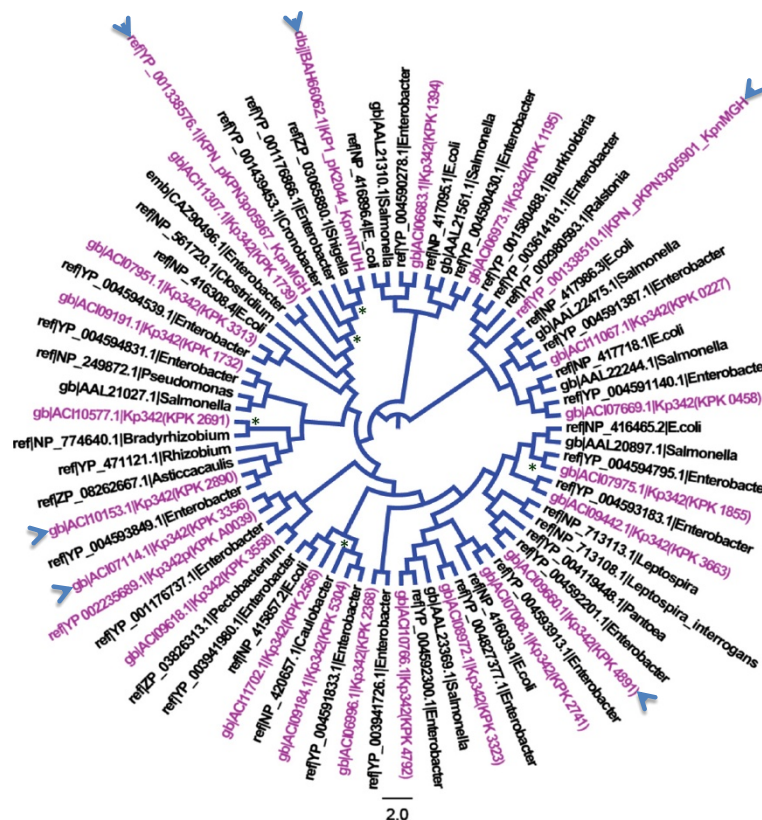


Figure 3 Phylogeny of *K. pneumoniae* GGDEF proteins. The phylogenetic reconstruction was done using neighbor-joining with 73 amino acid sequences from *K. pneumoniae* GGDEF proteins and other bacteria. Nodes with less than 70% support after 1000 bootstrap replicates are indicated with an asterisk. GGDEF proteins from Kp342, MGH78578 and NTUH-K2044 are highlighted in purple. Arrowheads represent the unique GGDEF proteins found in the *K. pneumoniae* strains (3 genomic and 3 plasmid-encoded genes). The scale bar indicates the number of amino acid substitutions per site.

C. crescentus and *P. aeruginosa*, no REC domains were identified. The phylogenetic analysis also indicated similarity with GGDEF proteins from other bacteria, which raises questions regarding the origin and distribution of these copies among multiple bacterial species. This analysis therefore shows parallels and differences with other bacteria and the presence of multiple proteins with diverse domain architecture that is indicative of a complex c-di-GMP network in *K. pneumoniae*. Future studies focused on the function of many of these DGC and PDE proteins might shed light on the processes involving growth and survival of this bacterium in different environmental settings.

Methods

The analysis was carried out with the following genomes: *K. pneumoniae* Kp342, *K. pneumoniae* MGH 78578 and *K. pneumoniae* NTUH-K2044 (GenBank NC_011283, NC_009648 and NC_012731, respectively). Genes coding for proteins with the GG(D/E)EF and E(A/V)L sequence motifs were identified with PSI-BLAST [38] using reference sequences available at NCBI Gene Entrez [39] [See Additional file 1], against the three *K. pneumoniae* genomes. Input sensory domains were identified using the databases CDD at the NCBI [40], InterproScan [41], pFam [42] and SMART [43]. Transmembrane segments were identified using SMART and SOSUIsignal [43,44], and the presence and localization of signal peptides was predicted using the SignalP 3.0 Server and SOSUIsignal [44,45]. Multiple alignments were done with the program MUSCLE [46] to identify the I site in each of the *K. pneumoniae* GGDEF domain proteins. Finally, the Genomic BLAST database from NCBI [38] was used to identify homologous GGDEF/EAL proteins in these three genomes. For all homologous proteins, Blastp was performed and the following parameters were considered: E-value less than 10^{-6} , identity percentage greater than 85% and query coverage greater than 95%. The homologous protein obtained was validated by Random Shuffling through PRSS/PREFX, using 500 shuffles [47]. The phylogenetic reconstruction was done with MEGA 5.05 [48], using 73 amino acid sequences and the neighbor-joining method with 1000 bootstrap replicates. Sequences from other families of Bacteria were selected from the Signaling Census database [20].

The logo sequences were generated using WebLogo 3.0 [49]. For DGCs we used an alignment of 9 DGC sequences [GenBank: YP_653766.1, YP_002517919.1, YP_258266.1, NP_252391.1, YP_631414.1, YP_471572.1, NP_459380.1, NP_463410.1, NP_416465.2] and 40 *K. pneumoniae* single-domain DGCs identified here. The logo for the PDE domain was done from an alignment of 7 PDE sequences [GenBank: AAC23902.1, AAC76550.2, ABJ13888.1, AAG07334.1, ACP09769.1, AAC73418.1,

CAB13282.1] and 40 *K. pneumoniae* PDEs identified here. The alignments were done using MUSCLE [46].

Additional files

Additional file 1: Title: Inventory of GGDEF proteins in *K. pneumoniae* 342, MGH 78578 and NTUH-K2044.

Additional file 2: Title: Inventory of EAL proteins in *K. pneumoniae* 342, MGH 78578 and NTUH-K2044.

Abbreviations

DGC, diguanylate cyclase; PDE, phosphodiesterase; c-di-GMP, Bis-(3'-5')-cyclic dimeric GMP; REC, receiving domain with phosphoacceptor site; CACHE, Ca²⁺ channels and chemotaxis receptors domain; CHASE, cyclase/histidine kinases-associated sensing extracellular domain; MASE, Membrane-associated sensor domain; PAS, PER, ARNT and SIM domain; HAMP, Histidine kinases, Adenylyl cyclases, Methyl binding proteins, Phosphatases domain; GAF, cGMP phosphodiesterase, adenylyl cyclase domain; BLUF, Sensing blue-light using FAD.

Competing interests

The authors declare that they have no competing interest.

Authors' contribution

The bioinformatics analysis was carried out by DC, analysis of results and discussions were done by DC, MH, ML, LZ and MMZ, the manuscript was prepared by DC, MH, ML, LZ and MMZ. All authors read and approved the final manuscript.

Acknowledgements

The work was financed by Colciencias (project No. 657045921709). We would like to thank J.M. Anzola, D. Riaño, J. Rodríguez and D. Chaves for discussions and help with the bioinformatics analysis.

Author details

¹Molecular Genetics, Corporación Corpogen, Carrera 5 No. 66A-34, Bogotá, Colombia. ²Department of Biological Sciences, Universidad de los Andes, Bogotá, Colombia. ³Department of Biological Sciences, Pontificia Universidad Javeriana, Bogotá, Colombia.

Received: 19 June 2012 Accepted: 25 June 2012

Published: 9 July 2012

References

1. Hoyos-Orrero SR-RO, Hoyos-Posada C, Mesa-Restrepo C, Alfaro-Velásquez M: Características clínicas, epidemiológicas y de susceptibilidad a los antibióticos en casos de bacteriemia por *Klebsiella pneumoniae* en neonatos. *Rev CES Med* 2007, **21**(2):31–39.
2. Struve C, Krogfelt KA: Pathogenic potential of environmental *Klebsiella pneumoniae* isolates. *Environ Microbiol* 2004, **6**(6):584–590.
3. Podschun R, Ullmann U: *Klebsiella* spp. as nosocomial pathogens: epidemiology, taxonomy, typing methods, and pathogenicity factors. *Clin Microbiol Rev* 1998, **11**(4):589–603.
4. Yu VL, Hansen DS, Ko WC, Sagnimeni A, Klugman KP, von Gottberg A, Goossens H, Wagener MM, Benedi VJ: Virulence characteristics of *Klebsiella* and clinical manifestations of *K. pneumoniae* bloodstream infections. *Emerg Infect Dis* 2007, **13**(7):986–993.
5. Marschall J, Fraser VJ, Doherty J, Warren DK: Between community and hospital: healthcare-associated gram-negative bacteremia among hospitalized patients. *Infect Control Hosp Epidemiol* 2009, **30**(11):1050–1056.
6. Fouts DE, Tyler HL, DeBoy RT, Daugherty S, Ren Q, Badger JH, Durkin AS, Huot H, Shrivastava S, Kothari S, et al: Complete genome sequence of the N₂-fixing broad host range endophyte *Klebsiella pneumoniae* 342 and virulence predictions verified in mice. *PLoS Genet* 2008, **4**(7):e1000141.
7. Balestrino D, Ghigo JM, Charbonnel N, Haagensen JA, Forestier C: The characterization of functions involved in the establishment and maturation of *Klebsiella pneumoniae* in vitro biofilm reveals dual roles for surface exopolysaccharides. *Environ Microbiol* 2008, **10**(3):685–701.

8. Boddicker JD, Anderson RA, Jagnow J, Clegg S: **Signature-tagged mutagenesis of *Klebsiella pneumoniae* to identify genes that influence biofilm formation on extracellular matrix material.** *Infect Immun* 2006, **74**(8):4590–4597.
9. Balestrino D, Haagensen JA, Rich C, Forestier C: **Characterization of type 2 quorum sensing in *Klebsiella pneumoniae* and relationship with biofilm formation.** *J Bacteriol* 2005, **187**(8):2870–2880.
10. Di Martino P, Cafferini N, Joly B, Darfeuille-Michaud A: ***Klebsiella pneumoniae* type 3 pili facilitate adherence and biofilm formation on abiotic surfaces.** *Res Microbiol* 2003, **154**(1):9–16.
11. Johnson JG, Clegg S: **Role of MrkJ, a phosphodiesterase, in type 3 fimbrial expression and biofilm formation in *Klebsiella pneumoniae*.** *J Bacteriol* 2010, **192**(15):3944–3950.
12. Langstraat J, Bohse M, Clegg S: **Type 3 fimbrial shaft (MrkA) of *Klebsiella pneumoniae*, but not the fimbrial adhesion (MrkD), facilitates biofilm formation.** *Infect Immun* 2001, **69**(9):5805–5812.
13. Barends TR, Hartmann E, Griese JJ, Beitlich T, Kirienko NV, Ryjenkov DA, Reinstein J, Shoeman RL, Gomelsky M, Schlichting I: **Structure and mechanism of a bacterial light-regulated cyclic nucleotide phosphodiesterase.** *Nature* 2009, **459**(7249):1015–1018.
14. Johnson JG, Murphy CN, Sippy J, Johnson TJ, Clegg S: **Type 3 fimbriae and biofilm formation are regulated by the transcriptional regulators MrkHI in *Klebsiella pneumoniae*.** *J Bacteriol* 2011, **193**(14):3453–3460.
15. Wilksch JJ, Yang J, Clements A, Gabbe JL, Short KR, Cao H, Cavaliere R, James CE, Whitchurch CB, Schembri MA, et al: **MrkH, a novel c-di-GMP-dependent transcriptional activator, controls *klebsiella pneumoniae* biofilm formation by regulating type 3 fimbriae expression.** *PLoS Pathog* 2011, **7**(8):e1002204.
16. Schirmer T, Jenal U: **Structural and mechanistic determinants of c-di-GMP signalling.** *Nat Rev Microbiol* 2009, **7**(10):724–735.
17. Hengge R: **Principles of c-di-GMP signalling in bacteria.** *Nat Rev Microbiol* 2009, **7**(4):263–273.
18. Cotter PA, Stibitz S: **c-di-GMP-mediated regulation of virulence and biofilm formation.** *Curr Opin Microbiol* 2007, **10**(1):17–23.
19. Wu KM, Li LH, Yan JJ, Tsao N, Liao TL, Tsai HC, Fung CP, Chen HJ, Liu YM, Wang JT, et al: **Genome sequencing and comparative analysis of *Klebsiella pneumoniae* NTUH-K2044, a strain causing liver abscess and meningitis.** *J Bacteriol* 2009, **191**(14):4492–4501.
20. Galperin MY: **Bacterial signal transduction network in a genomic perspective.** *Environ Microbiol* 2004, **6**(6):552–567.
21. Martoglio B, Dobberstein B: **Signal sequences: more than just greasy peptides.** *Trends Cell Biol* 1998, **8**(10):410–415.
22. Walter P, Johnson AE: **Signal sequence recognition and protein targeting to the endoplasmic reticulum membrane.** *Annu Rev Cell Biol* 1994, **10**:87–119.
23. Galperin MY, Nikolskaya AN, Koonin EV: **Novel domains of the prokaryotic two-component signal transduction systems.** *FEMS Microbiol Lett* 2001, **203**(1):11–21.
24. Tamayo R, Pratt JT, Camilli A: **Roles of cyclic diguanylate in the regulation of bacterial pathogenesis.** *Annu Rev Microbiol* 2007, **61**:131–148.
25. Mascher T, Helmann JD, Unden G: **Stimulus perception in bacterial signal-transducing histidine kinases.** *Microbiol Mol Biol Rev* 2006, **70**(4):910–938.
26. Ryan RP, Fouhy Y, Lucey JF, Dow JM: **Cyclic di-GMP signaling in bacteria: recent advances and new puzzles.** *J Bacteriol* 2006, **188**(24):8327–8334.
27. Jenal U, Malone J: **Mechanisms of cyclic-di-GMP signaling in bacteria.** *Annu Rev Genet* 2006, **40**:385–407.
28. Ho YS, Burden LM, Hurley JH: **Structure of the GAF domain, a ubiquitous signaling motif and a new class of cyclic GMP receptor.** *EMBO J* 2000, **19**(20):5288–5299.
29. Pappalardo L, Janausch IG, Vijayan V, Zientz E, Junker J, Peti W, Zweckstetter M, Unden G, Griesinger C: **The NMR structure of the sensory domain of the membranous two-component fumarate sensor (histidine protein kinase) DcuS of *Escherichia coli*.** *J Biol Chem* 2003, **278**(40):39185–39188.
30. Zhulin IB, Nikolskaya AN, Galperin MY: **Common extracellular sensory domains in transmembrane receptors for diverse signal transduction pathways in bacteria and archaea.** *J Bacteriol* 2003, **185**(1):285–294.
31. Anantharaman V, Aravind L: **The CHASE domain: a predicted ligand-binding module in plant cytokinin receptors and other eukaryotic and bacterial receptors.** *Trends Biochem Sci* 2001, **26**(10):579–582.
32. Gomelsky M, Klug G: **BLUF: a novel FAD-binding domain involved in sensory transduction in microorganisms.** *Trends Biochem Sci* 2002, **27**(10):497–500.
33. Seshasayee AS, Fraser GM, Luscombe NM: **Comparative genomics of cyclic-di-GMP signalling in bacteria: post-translational regulation and catalytic activity.** *Nucleic Acids Res* 2010, **38**(18):5970–5981.
34. Newell PD, Monds RD, O'Toole GA: **LapD is a bis-(3',5')-cyclic dimeric GMP-binding protein that regulates surface attachment by *Pseudomonas fluorescens* Pf0-1.** *Proc Natl Acad Sci U S A* 2009, **106**(9):3461–3466.
35. Christen M, Christen B, Folcher M, Schauerer A, Jenal U: **Identification and characterization of a cyclic di-GMP-specific phosphodiesterase and its allosteric control by GTP.** *J Biol Chem* 2005, **280**(35):30829–30837.
36. Grant JR, Stothard P: **The CGView Server: a comparative genomics tool for circular genomes.** *Nucleic Acids Res* 2008, **36**(Web Server issue):W181–W184.
37. Dutta C, Pan A: **Horizontal gene transfer and bacterial diversity.** *J Biosci* 2002, **27**(1 Suppl 1):27–33.
38. Cummings L, Riley L, Souvorov A, Resenchuk S, Dondoshansky I, Tatusova T: **Genomic BLAST: custom-defined virtual databases for complete and unfinished genomes.** *FEMS Microbiol Lett* 2002, **216**(2):133–138.
39. Maglott D, Ostell J, Pruitt KD, Tatusova T: **Entrez Gene: gene-centered information at NCBI.** *Nucleic Acids Res* 2011, **39**(Database issue):D52–D57.
40. Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, et al: **CDD: a Conserved Domain Database for the functional annotation of proteins.** *Nucleic Acids Res* 2011, **39**(Database issue):D225–D229.
41. Zdobnov EM, Apweiler R: **InterProScan—an integration platform for the signature-recognition methods in InterPro.** *Bioinformatics* 2001, **17**(9):847–848.
42. Finn RD, Mistry J, Tate J, Coggill P, Heger A, Pollington JE, Gavin OL, Gunasekaran P, Ceric G, Forslund K, et al: **The Pfam protein families database.** *Nucleic Acids Res* 2010, **38**(Database issue):D211–D222.
43. Schultz J, Milpetz F, Bork P, Ponting CP: **SMART, a simple modular architecture research tool: identification of signaling domains.** *Proc Natl Acad Sci U S A* 1998, **95**(11):5857–5864.
44. Gomi MSM, Mitaku S: **High performance system for signal peptide prediction: SOSUsignal.** *Chem-Bio Informatics Journal* 2004, **4**(4):142–147.
45. Petersen TN, Brunak S, von Heijne G, Nielsen H: **SignalP 4.0: discriminating signal peptides from transmembrane regions.** *Nat Methods* 2011, **8**(10):785–786.
46. Edgar RC: **MUSCLE: a multiple sequence alignment method with reduced time and space complexity.** *BMC Bioinforma* 2004, **5**:113.
47. Pearson WR: **Effective protein sequence comparison.** *Methods Enzymol* 1996, **266**:227–258.
48. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: **MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods.** *Mol Biol Evol* 2011, **28**(10):2731–2739.
49. Crooks GE, Hon G, Chandonia JM, Brenner SE: **WebLogo: a sequence logo generator.** *Genome Res* 2004, **14**(6):1188–1190.

doi:10.1186/1471-2180-12-139

Cite this article as: Cruz et al.: Comparative analysis of diguanylate cyclase and phosphodiesterase genes in *Klebsiella pneumoniae*. *BMC Microbiology* 2012 **12**:139.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

