SigMol: repertoire of quorum sensing signaling molecules in prokaryotes

Akanksha Rajput, Karambir Kaur and Manoj Kumar

Bioinformatics Centre, Institute of Microbial Technology, Council of Scientific and Industrial Research (CSIR), Sector 39-A, Chandigarh-160036, India

Received July 31, 2015; Revised September 17, 2015; Accepted October 06, 2015

ABSTRACT

Quorum sensing is a widespread phenomenon in prokaryotes that helps them to communicate among themselves and with eukaryotes. It is driven through quorum sensing signaling molecules (QSSMs) in a density dependent manner that assists in numerous biological functions like biofilm formation, virulence factors secretion, swarming motility, bioluminescence, etc. Despite immense implications, dedicated resources of QSSMs are lacking. Therefore, we have developed SigMol (http://bioinfo.imtech.res.in/manojk/sigmol), a specialized repository of these molecules in prokaryotes. SigMol harbors information on QSSMs pertaining to different quorum sensing signaling systems namely acylhomoserine lactones (AHLs), diketopiperazines (DKPs), 4-hydroxy-2-alkylquinolines (HAQs), diffusible signal factors (DSFs), autoinducer-2 (AI-2) and others. Database contains 1382 entries of 182 unique signaling molecules from 215 organisms. It encompasses biological as well as chemical aspects of signaling molecules. Biological information includes genes, preliminary bioassays, identification assays and applications, while chemical detail comprises of IUPAC name, SMILES and structure. We have provided user-friendly browsing and searching facilities for easy data retrieval and comparison. We have gleaned information of diverse QSSMs reported in literature at a single platform ‘SigMol’. This comprehensive resource will assist the scientific community in understanding intraspecies, interspecies or interkingdom networking and further help to unfold different facets of quorum sensing and related therapeutics.

INTRODUCTION

Quorum sensing (QS) is a signaling mechanism by which bacteria communicate among themselves and with other organisms (1–3). Through QS, they sense and respond to environmental changes via signal transduction events using signaling molecules in a density dependent manner (4). In QS, signaling molecules (autoinducers) are secreted out of the cell and on attaining a certain threshold these are sensed by other cells present in their vicinity. It further activates cascade of signaling events resulting in the activation of QS genes (5). This phenomenon was firstly reported by Nealson et al., in two bioluminescent marine bacterial species i.e. Vibrio fischeri and Vibrio harveyi (6). However, the term QS was coined by Greenberg et al. (5). Later on, this mechanism was also discovered in Natronococcus occultus, an archaean species (7).

Quorums sensing signaling molecules (QSSMs) are broadly distributed into different signaling systems namely acyl homoserine lactones (AHLs), quorum sensing peptides (QSPs) diketopiperazines (DKPs), diffusible signal factors (DSFs), 4-hydroxy-2-alkylquinolines (HAQs), autoinducer-2 (AI-2), autoinducer-3 (AI-3) and others (8). Amongst all the known QS signaling systems, AHLs are the most prevalent molecules predominantly found in Gram-negative bacteria (9). AHLs have acyl side chain that varies from C4-C18 of homoserine lactone moiety which are usually straight and in some cases may have branched configuration (9–11). AHLs are synthesized by LuxI or its homologues utilizing S-adenosylmethionine (SAM) and acyl-acyl carrier protein (acyl–ACP) as substrates (1). These signals, in turn sensed by LuxR or its homologues proteins leads to the activation of various physiological functions (9). Of the other QS signaling systems DKPs (12), DSFs (13), HAQs (14), AI-3 (15) are reported in Gram-negative bacteria, while QSPs are majorly found in Gram-positive bacteria (15). Moreover, AI-2 system is reported in both (16,17).

QSSMs help prokaryotes to adapt in diverse environment through various biological processes. One such important aspect is the development and dispersion of biofilms to cope up with harsh conditions. Biofilm formation is widely reported in numerous bacterial species, e.g. Streptococcus mutans (18), Pseudomonas aeruginosa (19), Vibrio cholerae (20), etc., whereas biofilm dispersion in Staphylococcus aureus, Vibrio cholerae, Xanthomonas campestris and so forth (21). Similarly other process like release of virulence fac-
tors causes extensive damage to the host. This helps bacteria to escape from host immune response as described in *Staphylococcus* spp. (22), *Streptococcus* spp. (23), *Burkholderia cepacia* complex (24) and many more. Likewise other QS mediated processes viz. swarming motility, genetic competence, bioluminescence, etc. also assist bacteria in multifarious ways (1, 15, 25). Apart from signaling functions, some QSSMs are also involved in non-signaling events like iron chelation, and membrane modification with the help of 2-heptyl-3-hydroxy-4-quinolone (PQS) (26, 27). Additionally, they also possess antimicrobial properties as mediated by autoinducer (lantibiotics) like nisin and subtilin from *Lactobacillus lactis* and *Bacillus subtilis*, respectively (28).

Despite inevitable importance of QSSMs, this field is still computationally under explored. Only one depository ‘Quorumpeps’ is available for QSPs system with 231 QS peptides entries from 51 bacteria (15). Another is an algorithm ‘QSPpred’ for analyzing and predicting QSPs (29). Therefore, there is an exigent need of a resource for majority of QS signaling systems. To fill this void, we have developed a comprehensive database ‘SigMol’ of QS signaling molecules in prokaryotes.

**MATERIALS AND METHODS**

**Data acquisition**

Exhaustive search of the literature was carried out to fetch relevant articles from PubMed. For this, keywords like quorum sensing and various signaling systems were used to build the final query as follows:

ˈquorum sensing) AND ((((((((((((acyl homoserine lactone) OR acyl-homoserine lactone) OR acylhomoserine lactone) OR acyl homoserine lactones) OR acyl homoserine lactones) OR acyl-homoserine lactones))))) OR (((DSF or Diffusible signal factor))) OR (((DKP or Diketopiperazines))) OR (((2-heptyl-3-hydroxy-4-quinolone or HHQ or pseudomonas quinolone signal or PQS))) OR (((AI*2) OR autoinducer*2))) OR (((AI*2) OR autoinducer*2))) OR autoinducer*2)

With this search query ∼2900 articles were obtained till August 2015. After initial screening ∼1400 potential articles were filtered to mine the relevant QSSMs information. Further reviews and articles lacking the required information were excluded and finally data was systematically extracted from 244 papers. In total, SigMol includes 1382 entries of 182 unique signaling molecules from 215 organisms.

We have provided chemical information of many QSSMs including their structure using chemical repositories viz., Pubchem or Chemspider. Structures of many signaling molecules were not found in these repositories therefore, we made their SMILES (Simplified molecular-input line-entry system) using Marvin sketch software (https://www.chemaxon.com/products/marvin/marvinsketch/) or Optical Structure Recognition (OSRA) (http://cactus.nci.nih.gov/cgi-bin/osra/index.cgi). Further, these SMILES were used to generate chemical information by utilizing Chemicalize.org (http://www.chemicalize.org/).

**Database organization and Web interface implementation**

SigMol is a specialized resource that will assist users to explore different QS molecules efficiently and effectively. It comprises information of diverse QSSMs manually curated and extracted from literature. Each entry in the database has been given a unique QSSM ID and contains fields like: (i) signaling system, (ii) signaling molecule, (iii) synthase and recipient genes, (iv) organism, (v) strain, (vi) PMID, etc. Individual entry of signaling molecule contains two types of information, i.e. chemical and biological. Chemical information includes name, abbreviation, IUPAC name, SMILES, molecular formula, molecular weight and structure. Whereas, biological information describes name of QSSM in article, synthase gene, recipient gene, organism along with its strain, preliminary bioassay, bacterial strain used in preliminary bioassay, identification assay, applications and article reference.

A unique feature of SigMol is the availability of drawings/structures of all the signaling molecules displayed under ‘Unique QSSMs’ menu on the webserver. It harbors signaling molecule, signaling system, structure of that molecule and link for extracting chemical information of the same. This resource is constructed using apache server on linux operating system. Database back-end is managed using an open source MySQL (relational database), Java script, PERL, HTML and PHP are used to develop front-end of the webserver. SigMol architecture is shown in Figure 1.

**RESULTS**

**Database statistics**

SigMol currently encompasses 1382 entries of 182 unique signaling molecules. Data curation helped to notify that different signaling systems, reported in literature till now are: AHLs, DKPs, DSFs, HAQs, AI-2, AI-3 and others. Out of all the signaling systems discussed earlier, AHLs correspond to maximum signaling molecules covering 985 entries. Here, N-hexanoyl-l-homoserine lactone (C6-HSL) has 156 entries followed by N-(3-oxohexanoyl)-l-homoserine lactone with 141 entries. Similarly, AI-2, HAQs, DSFs, DKPs contain 125, 120, 52, 48 signaling molecules, respectively. Remaining QSSMs, i.e. AI-3, CAI-1, α-pyrene, CHD and DAR are classified into ‘Others’ category having 52 entries (Figure 2A). These signaling molecules belong to 215 organisms including bacteria and archaea. For example, *Pseudomonas aeruginosa*, *Burkholderia pseudomallei*, *Entero bacter sakazakii*, *Aeromonas hydrophila* and *Yersinia ruckeri* are producing maximum number of QSSMs with 96, 72, 62, 54 and 52 entries, respectively (Figure 2B). Moreover, notable studies included in SigMol are from coral associated *Vibrios* (30), opportunistic pathogen (*Entero bacter sakazakii*) (31), food borne *Aeromonas* isolates (32), *Burkholderia* cluster (33), soft-rot bacteria (34) and many more.

As reported in the literature, QS phenomenon is driven mainly by two major genes viz. synthase gene to produce QSSM and recipient gene to sense respective signaling molecules. Major synthase genes presented in database are depicted in Figure 3A, among them luxI, luxS and
Figure 1. Architecture of SigMol.

Figure 2. Statistical distribution of quorum sensing signaling molecules (QSSMs) among prokaryotes: (A) signaling systems; (B) top 10 organisms producing them. [Abbreviations used: AHLs, acylated homoserine lactones; AI-2, autoinducer-2; DKPs, diketopiperazines; DSFs, diffusible signal factors; HAQs, 4-hydroxy-2-alkylquinolines].

*pqsH* produce 330, 122 and 76 QSSMs correspondingly in prokaryotes. Top 10 QSSM sensing genes are depicted in Figure 3B, out of which, *luxR*, *pqsR* and *ahyR* are reported to be present in maximum organisms’ viz. 328, 76 and 51, respectively.

**Data retrieval**

**Browse.** SigMol has been implemented with easy browsing facility. Users can browse the database by different browsing options or fields like signaling systems, genes and organisms. Further, browsing is divided in two-tiers that are based on signaling systems and individual signaling molecules (Supplementary Figure S1). User can choose required molecules for further details. Similarly, genes are also categorized in two parts viz. synthase gene and recipient gene. Simultaneously, organisms are grouped into two categories, i.e. alphabetically and in taxonomical order. Using these options users can browse the database in an easy and interactive way.
Search. In search option, query box is provided in which user can enter the query on the basis of different fields. Search type options include 'containing' and 'exact' facility. The search using 'containing' gives the output with the field containing entered keyword whereas 'exact' allows strict search. Output displays information, i.e. QSSM ID, signaling system, signaling molecule, synthase gene, recipient gene, identification assay and PMIDs of that particular query. Clicking individual QSSM ID displays detailed chemical, structural and biological information (Supplementary Figure S2).

Further, database is also hyperlinked to various external resources like PubChem (35), Chemspider (http://www.chemspider.com/), Chemicalize.org for extraction of chemical information. Genes in the database are linked to European Nucleotide Archive (ENA) and UniProt for additional details of DNA and protein sequences respectively. Further, organisms are linked to NCBI taxonomy browser. Each record in the resource is linked to PMID for meta information.

Tools. We have implemented two search tools to explore QSSMs namely 'compare' and 'draw' under 'tools' menu. Using first tool, user can select desired QS molecules from any signaling system to easily compare and visualize. Wherein, second tool allows the user to draw structure of a particular signaling molecule to search in the database. Apart from these tools, we have also provided links to QS related resources and metabolic pathways. 'Help' page will assist users to navigate SigMol web interface with ease.

Signaling molecules in intraspecies, interspecies and interkingdom communication. QSSMs are involved in intraspecies, interspecies and interkingdom communications as summarized in heatmaps (http://bioinfo.imtech.res.in/manoj/sigmol/summary.php). Intraspecies signaling occurs when a QSSM communicates within same species. For example, 10 Aeromonas and Vibrios, 9 Burkholderia spp. produce C6-HSL; similarly, C8-HSL is reported in 12 Burkholderia and 10 Vibrios indicate intraspecies communication among respective species (Figure 4).

However, presence of same QSSM in different prokaryotes reflects interspecies crosstalk clearly evident from Figure 4. For example, 27 different bacterial species (73 bacteria) reported to produce same signaling molecule C6-HSL, which is present in 10 Aeromonas and Vibrios; 9 Burkholderia spp; 5 Pseudomonas spp; 4 Dickeya, Serratia spp, and Yersinia spp; 3 Erwinia and Halomonas; 2 Chromobacterium spp, Pantoaea and Pectobacterium spp; and 15 other bacterial species. Likewise, C8-HSL, OC6-HSL, and C4-HSL are found in 24, 19 and 9 different bacterial species respectively (Figure 4).

Moreover, interkingdom communication is driven by a specific QSSM among organisms of different kingdoms (e.g. prokaryotes and eukaryotes). AI-3 is reported to be involved in interkingdom networking. For example, AI-3 helps bacterial species (Escherichia coli and Salmonella serovar Typhimurium) to crosstalk with human epinephrine/norepinephrine hormone during infection (3,36). Although limited, these QSSMs are also integrated in SigMol.

DISCUSSION

Presence of QS phenomenon among prokaryotes in regulating numerous physiological processes and aiding in crosstalk with eukaryotes further highlights its importance (4). In this study, we are providing a compendium ‘SigMol’, which integrates QSSMs of various QS signaling systems reported in prokaryotes since 1970. Inference from the data statistics revealed that many organisms have more than one QS signaling system like in E. coli (AI-2 and AI-3), Vibrio spp. (AHLs and AI-2), Dickeya spp. (AHLs and AI-2), etc. Amongst various bacteria provided in the repository, only Burkholderia spp. and Pseudomonas spp. showed presence of four QS systems namely AHLs, DPKs, DSFs and HAQs as depicted through heatmap at http://bioinfo.imtech.res.in/manoj/sigmol/summary.php.

Concurrently, within the same QS signaling system, a specific bacterium also generates diverse signaling molecules of that class. For example, 17 diverse AHLs have been reported for Sinorhizobium meliloti. Similarly, Burkholderia phytofirmans, Burkholderia xenovorans, Roseovarius toleri, Pseudomonas aeruginosa are also known to produce 15, 11, 11, 10 QSSMs, respectively (Supplementary Figure S3). It seems that existence of so many AHLs within the same
Acylated homoserine lactones (AHLs) (top 10) distribution in prokaryotes to represent intra- and inter-species communication. [Abbreviation used: C6-HSL, \(N\)-hexanoyl-L-homoserine lactone; C8-HSL, \(N\)-octanoyl-L-homoserine lactone; OC6-HSL, \(N\)-(3-oxohexanoyl)-L-homoserine lactone; C10-HSL, \(N\)-decanoyl-L-homoserine lactone; C12-HSL, \(N\)-dodecanoyl-L-homoserine lactone; OC10-HSL, \(N\)-(3-oxodecanoyl)-L-homoserine lactone; C14-HSL, \(N\)-tetradecanoyl-L-homoserine lactone; OHC8-HSL, \(N\)-(3-hydroxyoctanoyl)-L-homoserine lactone].

bacteria may help it to respond in different environments, however, this observation requires experimental validation.

AHLs signaling system is the most abundant and important among prokaryotes. We have categorized AHLs into five groups according to acyl chain modifications namely saturated, unsaturated, carbonyl, hydroxy and alamine methyl ester. Out of these, majority of AHLs belong to saturated and carbonyl followed by hydroxyl group. C6-HSL and C8-HSL are preferred among saturated AHL molecules, while OC6-HSL and OC8-HSL are for carbonyl group. Similarly, for hydroxyl group OHC8-HSL and OHC10-HSL are favored. Unlike bacteria, archaea have uncommon AHLs i.e. carboxylated-HSLs.

Formation of biofilm is the representative outcome of intricate patterns of communication to enhance pathogenicity of bacteria. In a classical example, two bacterial species Pseudomonas aeruginosa and Burkholderia cepacia are known to reside together in a biofilm in lungs of cystic fibrosis patients reflecting interspecies, interspecies and interkingdom networking (37,38). Likewise, multispecies biofilms (39) formed by various bacterial species involved in chronic wounds, dental plaque, etc. exhibit complex networking among different organisms. This QS based group-behavior of microbes is termed as ‘sociomicrobiology’ (40).

SigMol is a comprehensive resource of signaling molecules providing their biological and chemical information. It integrates all the facilities to explore QSSMs for searching signaling molecule of particular bacteria, browsing or comparing capability for specific systems and signaling molecules, structure based search and summary of all the QS systems present till date in the form of heatmaps. Here, all the prokaryotic QSSMs are integrated on one platform that can accelerate the research in field of quorum quenching therapeutics, mechanics and sociomicrobiology. Researchers can explore the role of signaling molecules to understand complex pattern of communication networking.

**AVAILABILITY**

SigMol, a comprehensive repository of QSSM is freely available at: http://bioinfo.imtech.res.in/manojk/sigmol/. We will update the database on half/yearly basis to include new information on QSSMs.

**SUPPLEMENTARY DATA**

Supplementary Data are available at NAR Online.

**FUNDING**

Department of Biotechnology, Government of India [GAP0001]; Council of Scientific and Industrial Research (CSIR) [GENESIS-BSC0121]. Funding for open access charge: CSIR-Institute of Microbial Technology. Conflict of interest statement. None declared.