African Journal of Biotechnology Vol. 11(57), pp. 12040-12044, 17 July, 2012
Available online at http://www.academicjournals.org/AJB
DOI: 10.5897/AJB12.1213
ISSN 1684–5315 ©2012 Academic Journals

Full Length Research Paper

Sequence analysis of putative swrW gene required for surfactant serrawettin W1 production from *Serratia marcescens*

Monabel May N. Apao*, Franco G. Teves and Ma. Reina Suzette B. Madamba

Department of Biological Sciences, College of Science and Mathematics, MSU-Iligan Institute of Technology, 9200 Iligan City, Philippines.

Accepted 22 June, 2012

Serratia marcescens produces biosurfactant serrawettin, essential for its population migration behavior. Serrawettin W1 was revealed to be an antibiotic serratamolide that makes it significant for deoxyribonucleic acid (DNA) and protein sequence analysis. Four nucleotide and amino-acid sequences from local strains analyzed through bioinformatics showed high confidence prediction of serrawettin. Database comparison analysis resulted to high similarity of the nucleotide sequence to the swrW gene of 88 to 94%, the homologous protein sequence to the serrawettin W1 synthetase protein ranging from 85 to 89%, presence of condensation domain from the non-ribosomal peptide synthetase (NRPS) family that synthesize peptide antibiotics and strong relation to the predicted surfactin synthetase structure. Further protein analysis showed high identical multiple alignment having conserved regions and the predicted structure representation was identified as putative surfactin a synthetase c (srfa-c), a non-ribosomal peptide synthetase termination module with 100% confidence. These nucleotide and protein sequence analysis of the putative swrW gene provides vital information on the versatility of *S. marcescens* as a pathogen of diverse hosts and an impetus for further genetic manipulation for practical applications.

Key words: *swrW* gene, serrawettin W1, non-ribosomal peptide synthetase (NRPS).

INTRODUCTION

Serratia marcescens is a member of the Enterobacteriaceae found in a variety of ecological niches, including soil, water, air, plants and animals (Grimont and Grimont, 1992). It has the ability to survive and grow under extreme conditions, such as in disinfectants (Marrie and Costerton, 1981; Parment et al., 1986), antiseptics (Nakashima et al., 1987) and double distilled water (Szewzyk et al., 1993). *S. marcescens* is distinctive in its exolipid product, the biosurfactant serrawettin. Serrawettin W1, W2 and W3, produced by *S. marcescens* at 30°C, are surface-active cyclodepsi-peptides essential for the population migration behavior of *S. marcescens* (Matsuyama et al., 1989; O'Rear et al., 1992). Serrawettin is a wetting agent on various surfaces, enhancer of flagellum-independent expansion of bacterial population on agar medium and accelerator of swarming on semi-solid agar medium and an antibiotic.

Serrawettin W1 produced by many pigmented *S. marcescens* strains (Matsuyama et al., 1986, 1989) is biosynthesized through the non-ribosomal peptide synthetase (NRPS) system, regulated by auto-induction system and a product of the *pswP* gene. Recently, serrawettin W1 synthetase putative gene *swrW* was identified through genetic analysis of serrawettin-less mutants of *S. marcescens* 274. This putative serrawettin synthetase gene is uni-modular in contrast to multi-modular nature of NRPS.

Serrawettin W1 encoded by the *swrW* gene belongs to NRPS family encoding a candidate for biosurfactant production and at the same time recognized as an

*Corresponding author. E-mail: feminine_blossom@yahoo.com. Tel:/Fax: +063 223 2343.

Abbreviations: NRPS, Non-ribosomal peptide synthetase; **srfa-c,** surfactin a synthetase c.

antibiotic called serratamolide by Matsuyama et al. (1986, 1989). In spite of having such obvious advantages, the genetic mechanisms of the production of serrawettin have not been thoroughly studied. Thus, to have a better understanding of the biosynthesis and biological activities of serrawettin, isolation of the *swrW* gene from local strains of *S. marcescens*, sequence comparison with homologues from other *S. marcescens* isolates in databases, and phylogenetic analysis, are essential. This is the first report of the *swrW* gene sequence from Philippine isolates of *S. marcescens*.

MATERIALS AND METHODS

Four local strains of *S. marcescens* (B111, B112, B211 and B212) in the culture collection of the Molecular Biology and Biotechnology Research Laboratory of MSU-IIT, Iligan City, Philippines, were used in this research. Streak plate cultures on Luria-Bertani (LB) medium (1.0 g tryptone, 0.5 g yeast extract, 1.0 g NaCl and 15.0 g nutrient agar per 100 ml of deionized water) were incubated at 27°C for three days. Isolated colonies were used for sub-cultures and for inoculation on LB broth for genomic deoxyribonucleic acid (DNA) isolation.

Genomic DNA isolation and purification using the Wizard® genomic DNA purification kit (PROMEGA Corporation)

One milliliter of an overnight *S. marcescens* LB broth culture incubated with shaking at 37°C was pipetted to make replicates in 1.5 ml micro centrifuge tubes. Cells were harvested and pelletized by centrifugation at 13,000 to 16,000x g for 2 min while the supernatant was discarded. Nuclei lysis solution (600 µl) was added to the cell pellet to resuspend the cells. The cell suspension was then incubated at 80°C for 5 min to lyse the cells, cooled and treated with ribonuclease (RNAse) solution (3.0 µl) with incubation at 37°C for 2 h. The suspension was allowed to cool down at room temperature for cell protein removal. The tubes were vortexed at high speed for 20 s to mix the 200 µl of protein precipitation solution with the cell lysate and placed on ice for 5 min and centrifuged at 13,000x g for 3 min. The residual liquid of the supernatant with the precipitated proteins was discarded.

Precipitation of genomic DNA was done by pipetting out the supernatant to new microcentrifuge tube containing 600 µl isopropanol kept at room temperature. The mixture was gently inverted several times to precipitate the DNA appearing as thread-like strands. This was followed by centrifugation at 13,000x g for 2 min. After the supernatant was discarded, residual liquid was drained completely on clean absorbent paper. The DNA pellet was then washed with 600 µl of 70% ethanol by inverting the tube and centrifuged after at 13,000x g for 2 min. The ethanol was pipetted out and remaining liquid was again drained followed by air-drying of the DNA pellet for 10 min at room temperature. DNA was re-suspended by the addition of 100µl of DNA rehydration solution with incubation at 65°C for 1 h for complete dissolution. Purified genomic DNA was stored in the refrigerator at 4°C after checking using agarose (0.8%) gel electrophoresis.

PCR amplification

Amplification was done using GoTaq® PCR Core System I kit (PROMEGA Corporation). Primers designed for PCR amplification of the published *swrW* gene which encodes putative serrawettin W1 synthetase were SW2-F3 (5′-GCG ACA AAA GCA ATG ACA AA) and SW2-R3 (5′-GTC GGC GTA TTG TTC CAA CT) with GenBank accession AB193098.2 synthesized by Oligo Service, MACROGEN Inc, Korea. The reaction mixtures were placed in a PTC-100 Programmable Thermal Cycler (MJ Research, Inc.) with the following cycling conditions: initial denaturation at 94°C, 5 min; annealing at 55°C, 2 min; and extension 72°C, 3 min; 30 cycles of denaturation at 94°C, 45 s; annealing 55°C, 45 s; and extension at 72°C, 3 min; 1 cycle of final extension at 72°C, 10 min. Polymerase chain reaction (PCR) products were stored at 4°C. Presence of amplified DNA products was checked by agarose (0.8%) gel electrophoresis.

DNA sequencing and nucleotide and amino acid sequence analysis

Quadruplicate PCR samples were sent to MACROGEN, Korea, for DNA sequencing. Nucleotide sequence analysis was done using bioinformatics open software (NCBI, ExPASy Swiss Bioinformatics, ClustalW, Phyre2 and PHYLIP). BLAST (Basic Local Alignment Search Tool) of NCBI (Altschul et al., 1997) was used to identify significant homologues, similar nucleotides and protein sequences from the databases. The open reading frames of the four nucleotide sequences were determined and then translated to protein sequences using ExPASy translation tool.

All four protein sequences of the local samples together with the top three amino acid sequences with significant alignment match namely, serrawettin W1 synthetase (*Serratia marcescens*) accession number ABL61529.1, putative serrawettin W1 synthetase (*Serratia marcescens*) accession number BAD60917.1, and amino acid adenylation domain-containing protein (*Serratia* sp. AS12) accession number YP004502934.1, were chosen for multiple alignment.

The four amino acid sequences with the top three significant sequence matches in the database were aligned using ClustalW. In PHYLogeny Inference Package (PHYLIP), the cladogram tree is constructed by ignoring the gaps and using the neighbor-joining (NJ) method. For further protein structure and function prediction, all seven protein sequences were submitted to protein homology/analogy recognition engine V 2.0 (Phyre2) to interpret protein structure and disorder prediction and alignment views (Kelley and Sternberg, 2009).

RESULTS AND DISCUSSION

DNA sequencing and nucleotide sequence analysis

Bioinformatics analysis showed that the four nucleotide sequences of the putative *swrW* gene in comparison with the database search revealed the same three highly similar sequence match from *S. marcescens* namely: *swrW* gene for putative serrawettin W1 synthetase, 4476 bp (AB193098.2), strain N4-5 serrawettin W1 synthetase gene, 748 bp (EF1220474.1), and strain ATCC 274 serrawettin W1 synthetase gene, 750 bp (EF1220477.1). *S. marcescens* strain ATCC 274 serrawettin W1 synthetase gene sequence match showed two separate locations of nucleotide sequence.

Sequence B111 having 915 bp is 94% (784/831) identical to *swrW* gene for putative serrawettin W1 synthetase (4476 bp) in the database with an E-value of 0.0. Sequence B112 having 952 bp is 94% (776/822) identical to the same gene with an E-value of 0.0. While

sequence B211 having 975 bp is 93% (804/860) identical and B212, 925 bp is 88% (592/676) identical to the gene with an E-value of 0.0. The other two match sequences (*Serratia marcescens* strain N4-5 serrawettin W1 synthetase gene, and *Serratia marcescens* strain ATCC 274 serrawettin W1 synthetase gene) in the database also showed high similarity percentage of 87 to 99%. Therefore, BLASTn confirmed that all four query sequences were homologous and highly similar to the *swrW* gene of *Serratia marcescens* strains in the database.

Amino acid sequence analysis

All four protein sequences of the *swrW* gene exhibited the same three significant protein sequence match from *S. marcescens* in the database search namely: serrawettin W1 synthetase, 211aa (ABL61529.1), putative serrawettin W1 synthetase, 1310 aa (BAD60917.1) and amino acid adenylation domain-containing protein, 1313aa (YP_004502934.1). Statistically, B111 having 259 aa is 89% (185/209) identical to serrawettin W1 synthetase (211aa) in the database with an E-value of 2e-131, 1 gap. Sequence B112 having 254 aa is 88% (164/187) identical to the same protein with an E-value of 7e-111, 2 gaps. While sequence B211 having 194 aa is 89% (87/98) identical with an E-value of 8e-57, no gaps and B212, 53aa is 85% (40/47) identical with an E-value of 9e-21, no gaps.

The other two match sequences (putative serrawettin W1 synthetase and amino acid adenylation domain-containing protein) in the database showed a percent similarity of 64 to 90%. Therefore, BLASTp confirmed that all four local amino acid sequences encodes putative serrawettin W1 synthetase protein indicating that the two sequences (query sequence and match sequence) evolved from a common ancestor and have the same function or structural organization. The result of BLASTp homology search confirmed that all four protein sequences have putative conserved domains indicating distinct functional and/or structural units of a protein.

Homology analysis of the 43 rows protein sequence alignment showed the presence of condensation domain, a characteristic of NRPS family. The condensation domain having an E-value of 1.61e-22 is found in many multi-domain enzymes which synthesize peptide antibiotics (Szewzyk et al., 1993). This domain catalyzes a condensation reaction to form peptide bonds in NRP biosynthesis. It is usually found to the carboxy side of a phosphopantetheine binding domain (pfam00550, p-p binding). It has been shown that mutations in the HHXXXDG motif abolish activity suggesting this is part of the active site.(Marchler-Bauer et al., 2004, 2009, 2011)

BLAST neighbor-joining tree showed the sequences are closely aligned and related to the Enterobacteria. Related protein structures result of BLAST and conserved domains search confirmed the protein sequences submitted where highly identical, aligned and related to the structure of the target surfactin a synthetase C (srfa-C), a non ribosomal peptide synthetase termination module (ligase) of *Bacillus subtilis* (Cosmina et al., 1993), where *B. subtilis* also produces prodigiosin. Moreover, *S. marcescens* have parallel production of serrawettiin and prodigiosin.

Multiple sequence alignment

The four local protein sequences and three database protein sequences in FASTA format were aligned using ClustalW2.1 multiple alignment (Figure 1). Alignment result of the seven protein sequences exhibited high level of conservation having four conserved positions on the entire column (represented by *), 10 residues of conserved substitutions and five columns of semi-conserved substitutions where the size or the hydropathy has been preserved in the course of evolution.

The tree result having high bootstrap values hypothesizes B112 (VIRT29469) as having the closest relationship to B212 (VIRT7608), B111 (VIRT24314) and gi|333929355-amino acid adenylation domain-containing protein, and distinct from gi|54114895-putative serrawettin W1 synthetase (outgroup). B211 (VIRT32126) and gi|119220920-serrawettin W1 synthetase form a sister clad indicating distantly related strains (Figure 2).

Protein structure prediction using protein homology/analogy recognition engine 2

All four protein sequences resulted to a putative srfa-c, a2 non ribosomal peptide synthetase termination module with 100% confidence by the single highest scoring template. Sequence B111 and B211 showed the highest percent coverage having 241 residues (96% of the sequence) and 187 residues (96% of the sequence) among the four sequences of the local *S. marcescens* strains. The three top sequences in the database also showed a 100% confidence of the 96% of the sequence as shown in Figure 3.

While amide synthase protein, tyrocidine synthetase a ligase and CoA-dependent acyltransferases belonging to NRPS condensation domain (amide synthase) are 84 to 87% confident match to the protein. This highly related domain match suggests that the protein of putative *swrW* strongly belongs to the NRPS family. Recent studies also revealed that homology analysis of *swrW* gene demonstrated the presence of condensation, adenylation, thiolation and thioesterase domains characteristic of NRPS.

Future studies will include; (a) gene knockout experiments to observe the effect of gene inactivation

```
VIRT29469                       FDSFLICGGGSVSAYSLTTVQQAVWLDQSLNPDIPLYNVGCLWRVEREIS 50
VIRT7608                        ----------------------------------------------SARFR 5
VIRT24314                       FDSFLICGGGSVSAYSLTTVQQAVWLDQSLNPDIPLYNVGCLWRVEREIS 50
gi|119220920|gb|ABL61529.1|     -----------MSAYSLTTVQQAVWLDQSLNPDIPLYNVGCLWRVEREIS 39
gi|54114895|dbj|BAD60917.1|     -----------MSAYSLTTVQQAVWLDQSLNPDIPLYNVGCLWRVEREIS 39
gi|333929355|ref|YP_004502934.  -----------MSAYSLTTVQQAVWLDQSLNPEIPLYNVGCLWHVEREIS 39
VIRT32126                       --------------------------------------------------

VIRT29469                       LPLFQEAIWQIQQQHDALQTTLKETPQGNTQETCDRAYIDLHYHDFSHRQ 100
VIRT7608                        IPLFQEAIRQIQQQHDALQTTLKETPQGVPQDTCDRAYISCLYHDFSH-- 53
VIRT24314                       LPLFQEAIRQIQQQHDALQTTLKETPQGITQETCDRAYIDRQYHDFSHRR 100
gi|119220920|gb|ABL61529.1|     LPLFQEAIRQIQQQHDALQTTLKETPQGITQETCDRAYIDLHYHDFSQRQ 89
gi|54114895|dbj|BAD60917.1|     LPLFQEAIRQIQQQHDALQTTLKETPQGITQETCDRAYIDLHYHDFSQRQ 89
gi|333929355|ref|YP_004502934.  LPLFKEAIRQIQQQHDALQTTLKEAPQGVIQEPLFKKQIELHYHDFSKEN 89
VIRT32126                       ------CVANHAERNAARNYPGNLRP-GLYRPAVPRFFAQTRCHQPGGNT 43
                                      .: :  ::: * : . :  * *  : .  :   .   *: .
```

Figure 1. Multiple sequence alignment of putative PCR-amplified serrawettin W1 synthetase and known homologous sequences. VIRT24314, B111; VIRT29469, B112; VIRT32126, B211; VIRT7608, B212; gi|119220920, serrawettin W1 synthetase; gi|54114895, putative serrawettin W1 synthetase; gi|333929355, amino acid adenylation domain-containing protein.

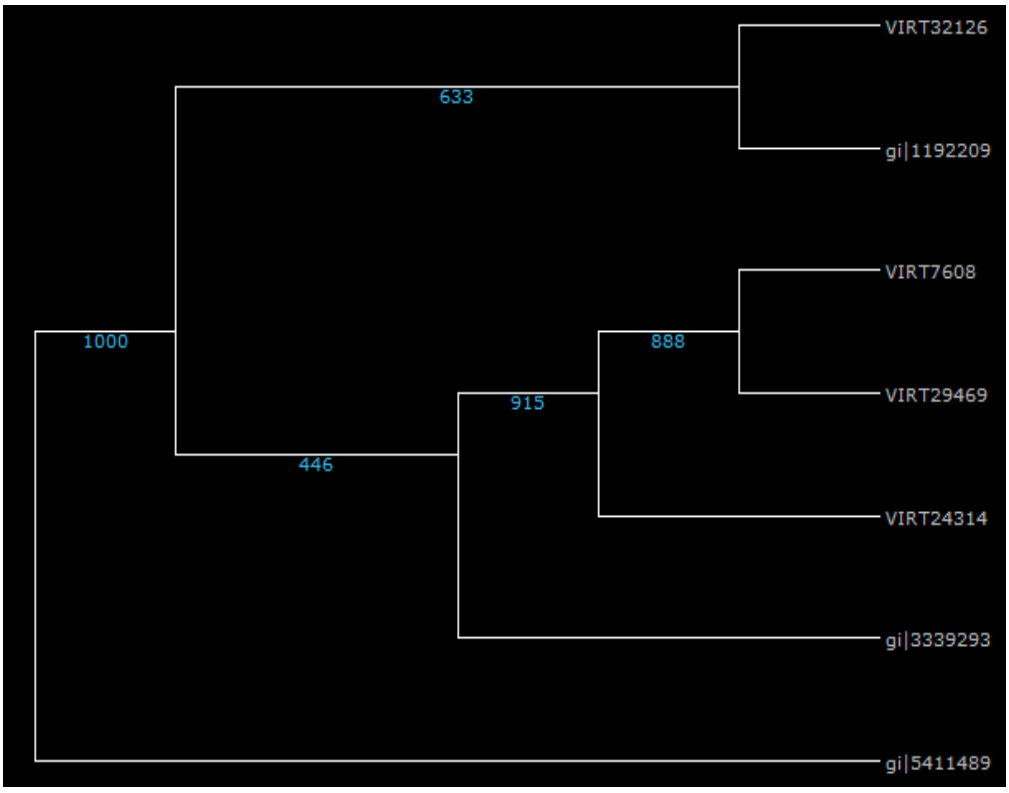

Figure 2. Cladogram tree by NJ method with bootstrap values above branches using PHYLIP program.

on serrawettin production; (b) serial deletions of the cloned target gene to determine the minimum sequence requirement for protein function; (c) complementation studies by transformation of serrawettin synthetase-deficient mutants with plasmid carrying the cloned putative serrawettin synthetase to observe restoration of serrawettin synthetase function; (d) western blot analysis to detect the presence of the serrawetin synthetase protein product in wild type and complemented mutant *S. marcescens* strains; and (e) toxicity and susceptibility

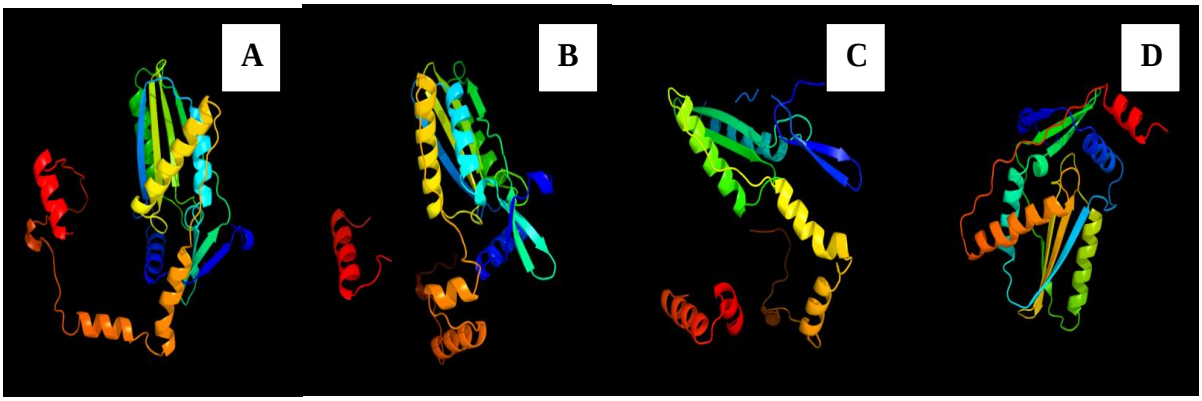

Figure 3. Four predicted protein structure- surfactin a synthetase c (srfa-c), a2 nonribosomal peptide synthetase termination module c2vsqA template model (Phyre2). A, B111; B, 112; C, B211; D, B212.

testing to determine the antibiotic and toxic effects of cell products from the different *S. marcescens* strains on test organisms. Such studies will also help to establish correlation, if any, between serrawettin production and pathogenicity.

Conclusion

Database searches and improved software tools are able to advance plausible predictions and accelerate research of the identification and characterization of the serrawettin gene from *S. marcescens*. Although the advanced predictions of the putative *swrW* gene appears very clear, experimental verification will have to be done to corroborate the *in silico* results. These are steps in designing genetic manipulations of the *swrW* gene for biosurfactant production and industrial synthesis of the antibiotic serratamolide, and understanding its role as a virulence factor in human infections.

ACKNOWLEDGEMENTS

This work is supported by the Department of Science and Technology-PCASTRD scholarship grant of the Philippine Government.

REFERENCES

Altschul SF, Madden TL, Schaffer AA, Zhang J, Miller W, Lipman DJ (1997). Gapped BLAST and PSI-BLAST:a new generation of protein database search programs. Nucleic Acids Res. 25:3389-3402.

Cosmina P, Rodriguez F, de Ferra F, Grandi G, Perego M, Venema G, van Sinderen D (1993). Sequence and analysis of the genetic locus responsible for surfactin synthesis in *Bacillus subtilis*. Mol. Microbiol. 8:821-831.

Grimont F, Grimont PAD (1992). The genus *Serratia*. In:Ballows A (Ed.), The Prokaryotes-A Handbook on the Biol. of Bacteria:Ecophysiol. Isolation, Identification, Applications, second ed., vol.III. Springer, New York, NY, pp. 2823-2848.

Kelley LA, Sternberg MJE (2009). Protein structure prediction on the web:a case study using the Phyre server Nature Protocols 4:363-371.

Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, Deweese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Lu F, Marchler GH, Mullokandov M, Omelchenko MV, Robertson CL, Song JS, Thanki N, Yamashita RA, Zhang D, Zhang N, Zheng C, Byant SH (additional)(2011). "CDD:a Conserved Domain Database for the functional annotation of proteins.", Nucleic Acids Res. 39(D):225-229.

Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, Deweese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Lu F, Marchler GH, Mullokandov M, Omelchenko MV, Robertson CL, Song JS, Thanki N, Yamashita RA, Zhang D, Zhang N, Zheng C, Byant SH (additional) (2009). "CDD:specific functional annotation with the Conserved Domain Database.", Nucleic Acids Res.37(D):205-210.

Marchler-Bauer A, Bryant SH (2004). *"CD*-Search:protein domain annotations on the fly.", Nucleic Acids Res.32(W)327-331

Marrie TJ, Costerton JW (1981). Prolonged survival of *Serratia marcescens* in chlorhexidine. Appl. Environ. Microbiol. 1981 Dec; 42(6):1093-102.PMID:7032422.

Marrie TJ, Costerton JW (1981). Prolonged survival of *Serratia marcescens* in chlorhexidine. Appl. Environ. Microbiol. 1981 Dec; 42(6):1093-102.PMID:7032422.

Matsuyama T, Murakami T, Fujita M, Fujita S, Yano I (1986). Extracellular vesicle formation and biosurfactant production by *Serratia marcescens*. J. General Microbiol. 132:865-875.

Matsuyama T, Sogawa M, Nakagawa Y (1989). Fractal spreading growth of *Serratia marcescens* which produces surface active exolipids. FEMS Microbiol. Lett. 61:243-246.

Nakashima AK, Highsmith AK, Martone J (1987). Survival of *Serratia marcescens* in benzalkonium chloride and in multiple-dose medication vials:relationship to epidemic septic arthritis s. J. Clin. Microbiol. 25.

O'Rear J, Alberti L, Harshey RM (1992). Mutations that impair swarming motility in *Serratia marcescens* 274 include but are not limited to those affecting chemotaxis or flagellar function. J. Bacteriol. 174:6125-6137.

Parment PA, Ronnerstamm R, Walder M (1986). Persistance of *Serratia marcescens, Serratia liquefaciens* and *E. coli* in solutions for contact lenses. Acta Ophthalmol (Co- penh) 64:456-462.

Zewzyk U, Szewzyk R, Stenstrom (1993). Growth and survival of *Serratia marcescens* under aerobic and anaerobic conditions in the presence of materials from blood bags. J. Clin. Microbiol. 31:1826-1830.