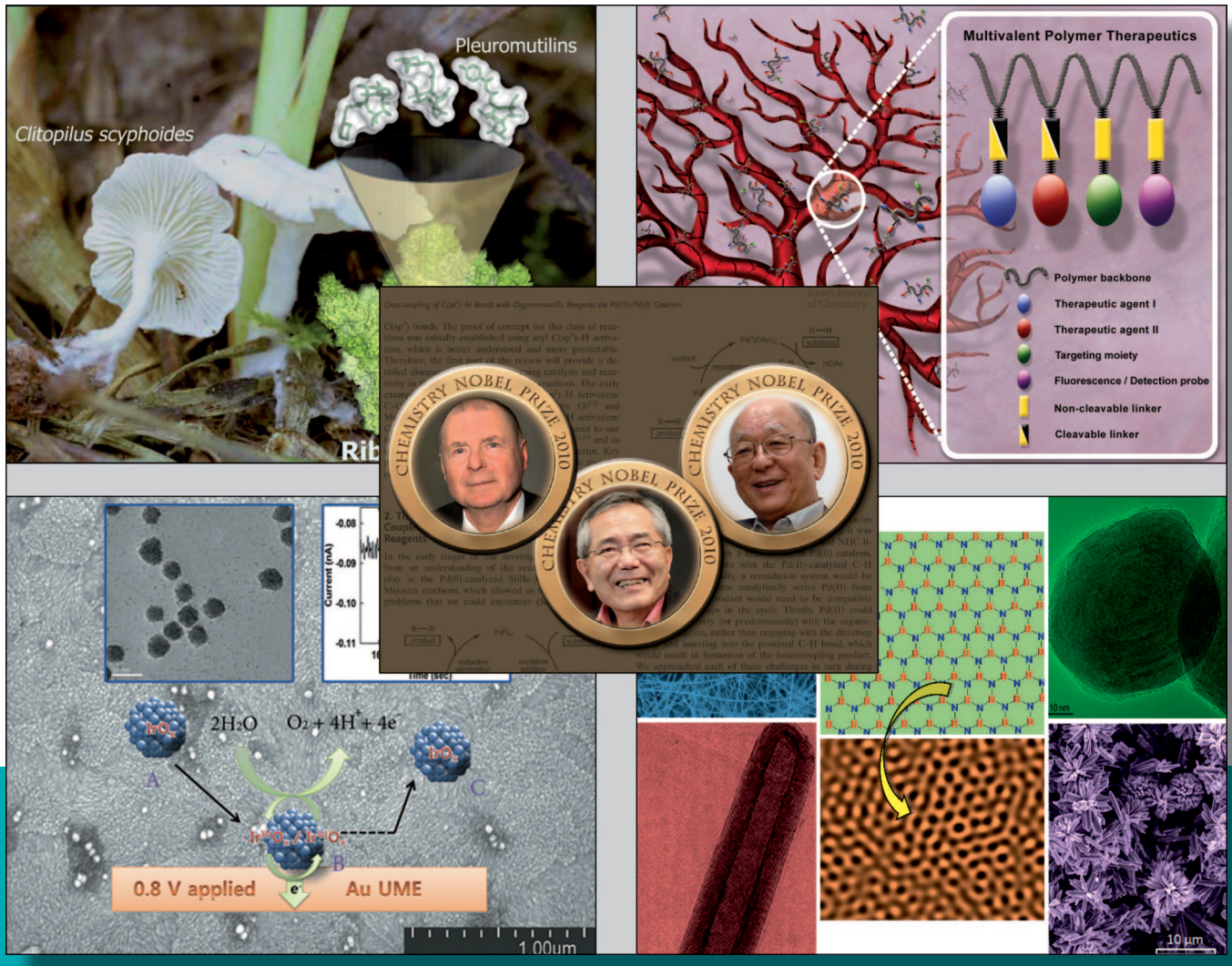
Israel Journal of Chemistry

IJC

Official Journal of the Israel Chemical Society

REPRINT

www.ijc.wiley-vch.de



WILEY-VCH

Communication

Israel Journal of Chemistry

DOI: 10.1002/ijch.201600133

Similarity between the Usher Plug and the Repeating Domain of an Ice-adhesin: Evolution via Surface Reshaping

Amit Kessel,[a] Rachel Kolodny,[b] and Nir Ben-Tal[a]

Abstract: The PapC usher and MpAFP ice-adhesin feature Ig-like domains, which are similar in shape and sequence but are engaged in very different functions. We explore how evolution reshaped the surfaces of these two domains to fit to their respective functions. In PapC, the Ig-like domain forms a rigid plug that seals the translocation channel in the inactive state. Upon activation, it undergoes a hinge motion, as an intact domain, to allow passage and assembly of the pili that goes through the pore. In accordance with this function, our calculations show that its surface is, in essence, electrostatically neutral, to facilitate the interaction with the PapC pore, and that this feature is evolutionarily conserved among its close homologues. On the other hand, within the context of MpAFP, the Ig-like domain is the main repeating unit, and our calculations show that its surface features a negative electrostatic potential of unusually high magnitude. The potential is balanced by bound Ca^{2+} ions. This is consistent with Ca^{2+}-dependent secretion of unfolded MpAFP across the double membrane of the bacterium, and folding outside, where Ca^{2+} concentration is high.

Keywords: evolution · electrostatics · unstructure · ConSurf · Ig-like domain

In proteins, evolutionary design from scratch (i.e., by adding one amino acid after another) is very challenging because the vast majority of such randomly generated amino acid sequences would not even fold properly. The alternative, more efficient means of evolution, is via co-option. That is, new proteins emerge by piecing together amino acid segments from existing proteins. With mutations, the old segments can be adjusted to their new environment and function. Here we describe one such anecdotal case, where a bacterial ice-adhesin and the PapC usher secretion system share an immunoglobulin (Ig)-like domain, which implies emergence from a common ancestor. Within the context of the approximately 0.7 μm long rod-like ice-adhesin, the Ig-like domain forms the main repeating element, which folds at high Ca^{2+} concentration. PapC, on the other hand, includes only a single copy of the domain, which serves as the channel plug. Evolutionary analysis of each protein within the contexts of its homologues shows that the surface of the plug is highly conserved. This makes sense because of the evolutionary pressure to retain its interaction with the channel pore; mutations might imped these functionally important interactions. In contrast, the surface of the adhesin repeating unit is evolutionarily variable, suggesting that the rod does not mediate highly specific interactions. The exception to this observation are the highly conserved Ca^{2+} binding residues. Indeed, adhesin folds only at high Ca^{2+} concentration, and the high conservation attests to the importance of Ca^{2+} binding at these sites. Our electrostatic analysis shows a remarkable difference between the two domains; the plug is practically neutral, whereas adhesin has a strong negative surface potential throughout, which is highly unusual. The high negative potential results from 19 acidic residues (and no basic residue) out of a total of 109 residues per domain. The high electrostatic repulsion between the numerous acidic residues within each domain, as well as between the domains, in particular those that are arranged in tandem, presumably unfolds the polypeptide chain. Ca^{2+} binding to their binding sites at the interfaces between the domains lowers the strong negative potential dramatically, and induces folding. The Ca^{2+} dependency of folding appears to be essential for secretion of adhesin from the cytoplasm to the outer environment of the cell through the narrow channel of Secretion System Type 1 (SST1). This demonstrates how evolution uses surface-redesign of a foldable segment, an Ig-like domain in this case, to facilitate a radical change of function.

The Ig-like fold is one of the most common protein domains.[1] As the name implies, it is particularly common in the immune system, where it comprises antibodies, receptors, co-receptors (e.g., CD4), and co-stimulatory proteins (e.g., CD28). It is also found in cell-adhesion proteins, as well as in various proteins of bacterial secretion systems. With such dominance, it is clear that the fold can adopt to various functions relatively easily. To examine the extent of this functional diversity, we inspect two appearances of the fold in very different bacterial proteins: An ice-adhesin protein and the PapC usher secretion system. With 16% sequence identity,

[a] A. Kessel, N. Ben-Tal
Department of Biochemistry and Molecular Biology, George S. Wise Faculty of Life Sciences, Tel Aviv University, Ramat Aviv, 69978, Israel
Homepage: URL: http://bental.tau.ac.il
[b] R. Kolodny
Department of Computer Sciences, University of Haifa, Mount Carmel 31905, Israel
Homepage: URL: http://cs.haifa.ac.il/~trachel

Israel Journal of Chemistry

(49% similarity) the two domains differ significantly from each other in their sequence. However, they superimpose with 5.9 Å RMSD. Furthermore, they share the same T class in ECOD (Immunoglobulin/Fibronectin type III/E set domains/PapD-like), which means that they feature the same topological connection, and are therefore considered to originate from a common ancestor.[2]

The ice-adhesin protein *Mp*AFP. MpAFP (*Marinomonas primoryensis* antifreeze protein) is a large (1.5 MDa) protein found in the Antarctic Gram-negative bacterium *Marinomonas primoryensis*. Because it can bind ice it was assumed to act as an intracellular antifreeze protein, hence the name.[3,4] However, later is was found that the protein resides outside the cell, that its ice-binding activity is located in a β-helix domain near the *C*-terminus, while it is bound to the cell via the *N*-terminus.[5] Thus, it is in fact an ice-adhesin, the vast majority of which is made of ~120 copies of a repetitive Ig-like domain of 109 amino acids.[6] Interestingly, this large repetitive region, called MpAFP RII (for region 2) folds only in the presence of high Ca^{2+} concentration.[7] It has been suggested that the Ca^{2+} dependency of folding is key to secretion of MpAFP from the cytoplasm across the two membranes of the bacterium.[6,7] According to this suggestion, MpAFP is translocated through SST1, which can transfer only unfolded proteins.[8] At the low Ca^{2+} concentration of the cytoplasm, MpAFP is unfolded, thereby capable of translocation through the narrow SST1 pore. The high Ca^{2+} concentration outside the cell below the layer of ice of a frozen lake that the bacterium inhabits, induces MpAFP folding and formation of a semi-rigid, rod-like structure that attaches the bacterium to the ice.[9]

The PapC usher. Many Gram-negative bacteria are coated with rod-like polymeric protein fibers called pili. These impotent elements attach the bacteria to their preferred niche. The pili is analogous to the ice-adhesin described above, however, while the ice-adhesin rod features repetitive element within a single polypeptide chain, the pili is made, for the most part, of many identical copies of one, Ig-like, protein called FimA for type I pili (and PapA for P pili). Because of the utmost importance of the pili to the bacterium, a special system, i.e., the usher protein, has evolved in order to transport the pili across the outer membrane of the bacterium. The usher, and a special chaperone, mediate the transport and assembly process of the pili.[8,10–14] The usher, is a kidney-shaped pore through the membrane with a β-barrel structure and an internal maximal cross section of 28×46 Å^2.[15,16] To prevent leakage across the outer membrane, the usher comprises an Ig-like domain, which serves as a plug. In the inactive state, the plug seals the usher pore (Figure 1). Upon delivery of the pili the plug swings away to allow transport.[15,17]

Evolutionary analysis. ConSurf calculations of the PapC usher and MpAFP ice adhesin show that the surfaces of their Ig-like domains (the plug and repetitive domains, respectively) differ significantly (Figure 2). The plug manifests a highly conserved surface. This is anticipated in view of its interaction with the PapC usher pore (Figure 1). On the other hand, the surface of the repetitive Ig-like domain of MpAFT is variable, which is typical of the protein exterior.[18] One of the regions in MpAFT that does exhibit high conservation is the interface between consecutive Ig-like domains (Figure 2b). Interestingly, the highly conserved residues in this region participate in coordination of two Ca^{2+} ions.

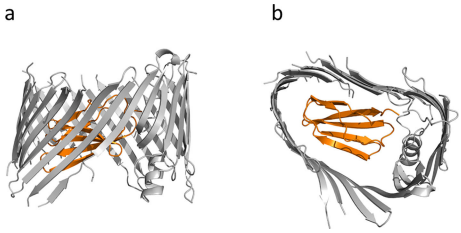


Figure 1. The PapC usher channel. Side (a) and top (b) views of the plugged channel (PDB entry 3fip). The plug domain is highlighted in orange.

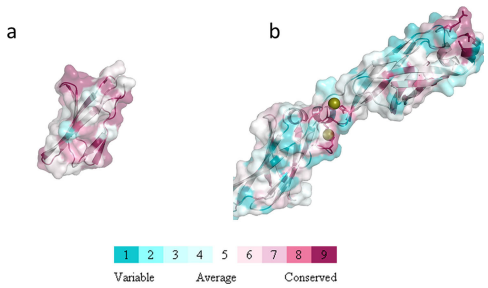


Figure 2. Evolutionary conservation analysis of the Ig-like domains of the PapC plug (a) and MpAFT repeat (b). Conservation was calculated using ConSurf,[24] with default setting [25] and the empirical Bayesian algorithm.[26] Each domain was taken within the contest of its respective intact protein. That is, PDB ID 4p99 for the MpAFT ice-adhesin, and 3fip for the PapC usher. The ConSurf grades, with conserved-through-variable corresponding to maroon-through-cyan, are mapped onto the (semi-transparent) protein surface. (b) For MpAFT two consecutive (identical) Ig-like domains are shown. The bound Ca^{2+} ions at the inter-domain interfaces are represented as spheres. The ions are coordinated by the highly conserved residues D224, N228, E231, D399, and N403 (grades 8–9), shown as sticks.

Un-structure prediction. Because of the observed Ca^{2+} dependency of MpAFT folding, we anticipated that its amino acid sequence composition may deviate from that of typical globular proteins, and will be predicted to be unstructured. To examine this hypothesis we used (with default setting) the IUpred web-server[19] (http://iupred.enzim.hu/), which is based on an estimated folding energy that depends only on the amino acid composition of the protein. MpAFT (UniProtKB: A0 A075B5G5) is predicted to be unstructured throughout; average un-structure probability of about 0.6 (values on a scale of 0–1, where 0 is fully structured and 1 is completely unstructured). In contrast, the PapC usher (UniProtKB: P07110) has only a short region (~residues 50–60) which is predicted to be unstructured (probability of about 0.6). The plug region is between amino acids 259–335, with average un-

Israel Journal of Chemistry

structure probability of 0.2. That is, the plug region is predicted to be structured, as expected.

Electrostatic analysis. To reveal the effect of Ca^{2+}-binding on MpAFT's structural stability we used the adaptive Poisson-Boltzmann solver (APBS)[20] to calculate the electrostatic potential of MpAFT with its bound Ca^{2+} ions. We then removed the ions and repeated the calculations. In the presence of Ca^{2+} ions MpAFT has a weak negative potential, with patches of positive potential in the domain interface regions (Figure 3a). In contrast, Ca^{2+}-free MpAFT is characterized by an excessively strong negative potential (Figure 3b), which results from the numerous acidic residues of the protein (19 per domain). The calculations may explain the effect of Ca^{2+} binding on MpAFT's structure. In the absence of Ca^{2+} the strong negative potential of MpAFT destabilizes the entire structure due to electrostatic repulsion between the acidic residues within each domain, as well as between consecutive domains. Such destabilization may prevent the protein from acquiring a folded structure. Ca^{2+} binding, due to the general Coulomb attraction, as well as stereo-specific effects (Figure 2b), may mask the acidic residues, reduce the electrostatic repulsion, and allows MpAFT to acquire its native fold. The rigidifying effect of Ca^{2+} binding has been noticed also for titin,[21,22] which is another protein composed of repeating Ig fold units. This protein provides muscle elasticity, and it has been suggested that Ca^{2+} binding modulates its function. It is noteworthy, however, that while the highly negative electrostatic potential of Ca^{2+}-free MpAFT decreases its structural stability, it is likely also to increase its solubility and confer resilience to aggregation (see also [23]). This property allows MpAFT to reversibly shift between the unfolded and the folded states without the risk of undergoing irreversible aggregation.

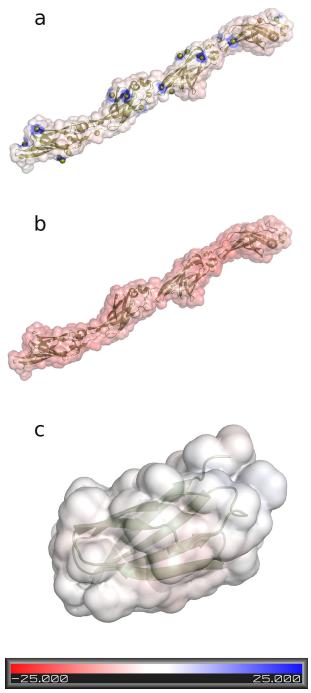


Figure 3. Electrostatic potential of the Ig-like domains of the PapC plug and MpAFT repeat. The potential is mapped onto the (semi-transparent) surface of the proteins, where negative potential ($0k_BT/e > \varphi > -25k_BT/e$) is red, positive potential ($0k_BT/e < \varphi < 25k_BT/e$) is blue, and neutral potential is white. Hydrogen atoms were added using the PDB2PQR server,[27] based on the ionization states calculated by PROPKA [28,29] at pH = 7. The electrostatic potential was calculated using APBS [20] with default parameters, and charges and radii taken from the PARSE forcefield.[30,31] (a) The potential of the Ig-like domain of MpAFT (PDB entry 4p99) when bound to Ca^{2+} (25 ions for the four tandem repeats). (b) The potential of the same domain in the absence of Ca^{2+}. (c) The potential of the PapC plug (PDB entry 3fip). This and the previous figures were constructed using PyMol.[32]

Unlike MpAFT, the PapC plug is folded in the absence of Ca^{2+}. To pinpoint the basis of this difference in stability, we calculated the electrostatic potential of the PapC plug domain. The plug has a near-neutral potential in the absence of Ca^{2+} ions (Figure 3c). Inspection of the amino acid composition of the two domains and their spatial distribution explains the differences in their potentials. First, the plug domain contains only 8 acidic residues, whereas the Ig-like domain of MpAFT has 19. Second, the plug domain also contains 7 basic residues (Arg, Lys), which mask the negative charge on some of the acidic residues. The Ig-like domain of MpAFT, on the other hand, does not contain any basic residues that can reduce its considerable negative charge.

By correlating the Ig-like domains of the PapC usher plug and MpAFP ice-adhesin using bioinformatics and electrostatics analyses, we show how evolution, via surface reshaping of the same Ig-like scaffold, can radically change function. These specific domains exist in different bacterial species, and tracing the exact evolutionary path between them, while interesting, is beyond the scope of this report.

Acknowledgements

This research is supported by grant No. 450/16 of The Israel Science Foundation to R. K. and N. B.-T.

References

[1] A. N. Barclay, *Semin. Immunol.* **2003**, *15*, 215–223.
[2] H. Cheng, R. D. Schaeffer, Y. Liao, L. N. Kinch, J. Pei, S. Shi, B. H. Kim, N. V. Grishin, *PLoS Comput. Biol.* **2014**, *10*, e1003926.
[3] J. A. Gilbert, P. L. Davies, J. Laybourn-Parry, *FEMS Microbiol. Lett.* **2005**, *245*, 67-72.
[4] C. P. Garnham, J. A. Gilbert, C. P. Hartman, R. L. Campbell, J. Laybourn-Parry, P. L. Davies, *Biochem. J.* **2008**, *411*, 171–180.
[5] S. Guo, C. P. Garnham, J. C. Whitney, L. A. Graham, P. L. Davies, *PLoS One* **2012**, *7*, e48805.
[6] T. D. Vance, L. L. Olijve, R. L. Campbell, I. K. Voets, P. L. Davies, S. Guo, *Biosci. Rep.* **2014**, *34*, 357–368.
[7] S. Guo, C. P. Garnham, S. Karunan Partha, R. L. Campbell, J. S. Allingham, P. L. Davies, *FEBS J.* **2013**, *280*, 5919–5932.

[8] T. R. Costa, C. Felisberto-Rodrigues, A. Meir, M. S. Prevost, A. Redzej, M. Trokter, G. Waksman, *Nat. Rev. Microbiol.* **2015**, *13*, 343-359.
[9] P. Delepelaire, *Biochim. Biophys. Acta* **2004**, *1694*, 149–161.
[10] H. Remaut, N. Ben-Tal, *eLife* **2014**, *3*, e04997.
[11] F. G. Sauer, K. Futterer, J. S. Pinkner, K. W. Dodson, S. J. Hultgren, G. Waksman, *Science* **1999**, *285*, 1058–1061.
[12] F. G. Sauer, J. S. Pinkner, G. Waksman, S. J. Hultgren, *Cell*, 111, 543–551.
[13] M. Nishiyama, T. Ishikawa, H. Rechsteiner, R. Glockshuber, *Science* **2008**, *320*, 376.
[14] G. Phan, H. Remaut, T. Wang, W. J. Allen, K. F. Pirker, A. Lebedev, N. S. Henderson, S. Geibel, E. Volkan, J. Yan, M. B. A. Kunze, J. S. Pinkner, B. Ford, C. W. M. Kay, H. Li, S. J. Hultgren, D. G. Thanassi, G. Waksman, *Nature* **2011**, *474*, 49–53.
[15] H. Remaut, C. Tang, N. S. Henderson, J. S. Pinkner, T. Wang, S. J. Hultgren, D. G. Thanassi, G. Waksman, H. Li, *Cell* **2008**, *133*, 640–652.
[16] Y. Huang, B. S. Smith, L. X. Chen, R. H. G. Baxter, J. Deisenhofer, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 7403–7407.
[17] I. Farabella, T. Pham, N. S. Henderson, S. Geibel, G. Phan, D. G. Thanassi, A. H. Delcour, G. Waksman, M. Topf, *Elife* **2014**, *3*.
[18] A. Kessel, N. Ben-Tal, in *Introduction to Proteins: Structure, Function, and Motion*, 1st ed., CRC Press, **2010**.
[19] Z. Dosztanyi, V. Csizmok, P. Tompa, I. Simon, *J. Mol. Biol.* **2005**, *347*, 827–839.
[20] N. A. Baker, D. Sept, S. Joseph, M. J. Holst, J. A. McCammon, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 10037–10041.
[21] H. L. Granzier, S. Labeit, *Circ. Res.* **2004**, *94*, 284–295.
[22] D. Labeit, K. Watanabe, C. Witt, H. Fujita, Y. Wu, S. Lahmers, T. Funck, S. Labeit, H. Granzier, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 13716–13721.
[23] M. S. Lawrence, K. J. Phillips, D. R. Liu, *J. Am. Chem. Soc.* **2007**, *129*, 10110–10112.
[24] F. Glaser, T. Pupko, I. Paz, R. E. Bell, D. Bechor-Shental, E. Martz, N. Ben-Tal, *Bioinformatics* **2003**, *19*, 163–164.
[25] H. Ashkenazy, S. Abadi, E. Martz, O. Chay, I. Mayrose, T. Pupko, N. Ben-Tal, *Nucleic Acids Res.* **2016**, *44*, W344–350.
[26] I. Mayrose, D. Graur, N. Ben-Tal, T. Pupko, *Mol. Biol. Evol.* **2004**, *21*, 1781–1791.
[27] T. J. Dolinsky, J. E. Nielsen, J. A. McCammon, N. A. Baker, *Nucleic Acids Res.* **2004**, *32*, W665–667.
[28] C. R. Søndergaard, M. H. M. Olsson, M. Rostkowski, J. H. Jensen, *J. Chem. Theory Comput.* **2011**, *7*, 2284–2295.
[29] M. H. M. Olsson, C. R. Søndergaard, M. Rostkowski, J. H. Jensen, *J. Chem. Theory Comput.* **2011**, *7*, 525–537.
[30] D. Sitkoff, K. A. Sharp, B. Honig, *J. Phys. Chem.* **1994**, *98*, 1978-1988.
[31] C. L. Tang, E. Alexov, A. M. Pyle, B. Honig, *J. Mol. Biol.* **2007**, *366*, 1475–1496.
[32] W. L. DeLano, DeLano Scientific LLC, San Carlos, CA, USA., 2002.

Received: October 14, 2016
Accepted: January 12, 2017
Published online on February 2, 2017