RecFinder – a new tool to find phage proteins responsible for host recognition

Angela O. Mucha, Franklin L. Nobrega, Silvio Santos, Eugénio C. Ferreira, Leon D. Kluskens, Joana Azeredo

IBB – Institute for Biotechnology and Bioengineering, Centre of Biological Engineering, Universidade do Minho, Campus de Gualtar, 4710-057 Braga, Portugal

Abstract

(Bacterio)phages are bacterial viruses that represent the most abundant and genetically diverse biological entities on the planet [1,2]. The largest virus group, the order Caudovirales (containing 96% of all known phages), has evolved recognition peptides for efficient virus-host interaction [3]. The peptides, usually located at the phage tail fiber, baseplate and other tail proteins mediate recognition and attachment specifically to bacterial cell wall receptors - lipopolysaccharides, teichoic acids, proteins and flagella. These proteins responsible for host recognition and binding (HBP) are macromolecular machines that make the process of infection highly efficient and finely regulated, and most likely play an important role in the evolutionary success of tailed phages. As a result of this specific binding affinity certain phages can only infect certain bacteria, determining in this manner the phage host range [4].

Over the last years, next-generation sequencing has led to an exponential increase of the number of completed phage genomes. Nevertheless, due to the poor knowledge on phage genetics and proteomics, a high percentage (60%-80%) of phage proteins is annotated as hypothetical protein. That percentage can increase in the case of HBP which present high variability. Therefore, HBP gene identification and functional analysis relies on slow and expensive laboratory work. This work is based on the *oligonucleotide-stickiness* method [5], where the *stickiness* (σ) – the binding analysis (or hybridization) stability of an oligonucleotide/peptide or protein – is used against a genome or proteome sequence of interest as follows:

$$\sigma = \frac{1}{n} \sum_{i=l_0+1}^{l_0+n} \delta(p, T(i)) \quad , \qquad (0 \leq \sigma \leq 1)$$

where $l0 + 1$ and n are the model positions at which the sampling region begins and the sampling size for *stickiness*, respectively, and δ is a determinant that takes the value of 1 when the sequence (p) binds stably to the i-th local sequence of the model (T (i)), that is, a fragmental sequence that has a fixed end at the sequence position i, or the value of 0 when there is no binding. In this formula, the p-T binding is determined based on the thermodynamic stability of the p-T complex under physiological conditions [5,6].

To prove the model, we choose T7 (a well-studied host-phage interaction model) HBP and an *Escherichia coli* model comprised of genes (genetic model) or proteins (proteomic model) present in the inner and outer membrane. Also, a study using PVP- SE1 phage and vB_CcoM-IBB_35 phage against, *Salmonella* and *Campylobacter* models, respectively, was performed. The proteomic vs. genomic approach is discussed. We hope that in the future this will lead to a

tool fed with all known sequenced bacteria (database) that will scan a certain sequence, from N- to C-terminal or 5- to 3-prime finding the specific binding domain in a hypothetical HBP.

References

1. Mahony, J., O. McAuliffe, *et al.* (2011). "Bacteriophages as biocontrol agents of food pathogens." Curr Opin Biotechnol 22(2): 157-163.
2. Mahichi, F., A. J. Synnott, *et al.* (2009). "Site-specific recombination of T2 phage using IP008 long tail fiber genes provides a targeted method for expanding host range while retaining lytic activity." FEMS Microbiol Lett 295(2): 211-217.
3. Schwarzer, D., K. Stummeyer, *et al.* (2007). "Characterization of a Novel Intramolecular Chaperone Domain Conserved in Endosialidases and Other Bacteriophage Tail Spike and Fiber Proteins." Journal of Biological Chemistry 282(5): 2821-2831.
4. Ackermann, H. W. 1998. "Tailed bacteriophages: the order caudovirales." Adv. Virus Res. 51:135-201.
5. Nishigaki, K. and A. Saito (2002). "Genome structures embossed by oligonucleotide- stickiness." Bioinformatics 18(9): 1153-1161.
6. Ahmed, S., A. Saito, *et al.* (2009). "Host–parasite relations of bacteria and phages can be unveiled by Oligostickiness, a measure of relaxed sequence similarity." Bioinformatics 25(5): 563-570.