# To know the ideal extremities of an EST

**Francisco Prosdocimi[1], Maurício A. Mudado[2], Fabiano C. Peixoto[3], J. Miguel Ortega[2]**

[1]Depto. Biologia Geral, [2]Depto. Bioquímica e Imunologia, [3]Laboratório de Computação Científica, Universidade Federal de Minas Gerais, 31270-010, MG, BRAZIL.

Expressed Sequence Tag is the result of a single pass sequence of a cloned cDNA molecule. ESTs are used to sample the presence of transcripts in transcriptomes represented by cDNA libraries. Usually, sequencing reaction is processed in automated equipments that produce a diagram of of consecutive pikes known as a chromatogram; the determination of the sequence of bases is obtained by base-caller software such as PHRED. We decided to generate in a single-pool sequencing reaction material enough for running around a thousand sequences of the same DNA molecule, pUC18. The resulting reads, not trimmed for low quality regions, were aligned to the published sequence of pUC18, thus determining the ideal position of trimming. Subsequently, PHRED trimming parameter trim_alt was calibrated varying the trim_cutoff setting to establish the ideal setting that would coincide with the results of alignment. We observed that admitting up to 10% error in such EST controlled approach, reads that loose information, i.e. the trimmed molecule is shorter than the ideal alignment, are in average 100 nt shorter and that reads that add undesired information, i.e. the trimmed molecule is longer than the ideal alignment, are not obtained. Interestingly, longer than desired reads is obtained only when the percentage of error admitted is over 16%, what corresponds to PHRED 8 value. Moreover, we demonstrated that reads trimmed with PHRED 8 bear 16% error in the extremity of the alignments, while carrying 3% global error, compatible with the goal of an EST. Translating the pUC18 sequence to a virtual protein, we were able to verify that BLASTx scores maximize if reads are trimmed with PHRED 8. We thus have proposed an experimental design to determine the correct far extremity of an EST. Using the same approach, we subsequently determined the ideal positioning of the sequencing primer, upstream of the vector/insert transition that would yield the transitions to appear in the non trimmed portion of the read. These experiments showed that positioning the primer at 60 bases from the transition, in average over 90% of the transitions will be seen, and 13 bases of vector would be collected in the reads [1]. Thus, we proposed an experimental determination of the near extremity of an EST. In our knowledge, this is the first experimental controlled determination designed to know the ideal extremities of an EST.

[1] Prosdocimi F, Ortega JM (2005). Accessing optimal primer distance from insert. In Silico Biol. 5 (5-6): 469-477.