
PREDICCIÓN DE SECUENCIAS REGULATORIAS DEL GEN *dlg* EN *Drosophila melanogaster* Y SU EVALUACIÓN IN VIVO MEDIANTE USO DE TRANSGÉNICOS

**ORLANDO ANTONIO CHÁVEZ ORREGO
INGENIERO EN BIOINFORMÁTICA**

RESUMEN

Un problema importante de la actual era de la genómica es identificar y caracterizar funcionalmente los genes relevantes para una vía o proceso biológico, para esto no sólo es necesario entender su función en un tejido sino también su regulación transcripcional.

El gen *dlg* de *Drosophila melanogaster* da origen a dos productos proteicos principales: DLGA, similar a la proteína neuronal de mamífero PDS95; y DLGS97, similar a la proteína neuronal y epitelial de mamífero SAP97. DLGA es expresado en epitelio, tejido neuronal y muscular mientras que DLGS97 es expresado en tejido neuronal y muscular, pero no en epitelio. En un orden temporal, DLGA se expresa de manera más marcada en estado de desarrollo temprano en tejido neuronal, a diferencia de DLGS97, el cual, aumenta sus niveles de expresión en este tejido en estado de mosca adulta. Estos dos transcritos se originan por inicios de transcripción alternativos lo que sugiere una regulación compleja donde múltiples Factores de Transcripción (FT) podrían estar participando. El objetivo de esta investigación es la identificación de la arquitectura de los promotores y los módulos regulatorios del gen *dlg*. Se pondrá énfasis en la identificación de los módulos de regulación que podrían explicar la expresión del gen neuronal. Un módulo de regulación es una región de ADN en la cual se unen múltiples FT y su influencia combinada produce un patrón de expresión específico. La hipótesis es que estos módulos se encuentran en regiones cercanas a los inicios de transcripción. El objetivo del trabajo es identificar los módulos (conjunto de motivos) que caracteriza cada uno de los patrones de expresión. Para ésta identificación se realizó un enfoque bioinformático de varios pasos que comienza con la comparación genómica de la secuencia de *dlg* de doce ortólogos disponibles de la familia *Drosophilidae* para seleccionar los motivos conservados. La selección de motivos conservados se comparó con bases de datos de Sitios de Unión (SU) de FT conocidos. Después de caracterizar los supuestos módulos de regulación del gen *dlg* se procedió a la búsqueda en todo el genoma para

identificar genes que comparten una regulación similar, en particular la búsqueda de genes de expresión neuronal. Finalmente, con las predicciones in-silico se crearon vectores experimentales con los putativos módulos río arriba y la secuencia nuclear GFP con la finalidad de probar in vivo la capacidad del módulo para conducir la expresión en un tejido específico.

Se determinaron 4 módulos que se probaron de forma independiente experimentalmente. Dos módulos estarían regulando DLGS97 y los otros dos a DLGA. Los módulos 1 y 2 en embriones transgénicos de *Drosophila melanogaster* estarían regulando la expresión de DLGA, y los módulos 3 y 4 la expresión de DLGS97. En larvas transgénicas de *Drosophila melanogaster* los módulos 1 y 4 estarían regulando la expresión de DLGA. En adultos transgénicos los módulos 1, 2 y 4 lo harían para la isoforma DLGS97. Las conclusiones se obtuvieron después de disectar cada uno de los estados de desarrollo de *Drosophila melanogaster* y ser observados en microscopios de fluorescencia.

ABSTRACT

An important subject in the current genomics era is to identify and functionally characterize relevant genes for a specific pathway or biological process, for this is not necessary to understand their function in a tissue but also its transcriptional regulation.

The *dlg* gene of *Drosophila melanogaster* gives rise to two major protein products: DLGA, similar to neuronal protein of mammal PDS95, and DLGS97, similar to neuronal and epithelial protein of mammal SAP97. DLGA is expressed in epithelial, neuronal and muscle tissues while DLGS97 is expressed in neuronal and muscle tissue but not epithelial. In a temporary order, DLGA is expressed more strongly in early development in neural tissue, unlike DLGS97, which increases their levels of expression in this tissue in the adult fly. These two transcripts originate from alternative transcription starts suggesting a complex regulation where multiple transcription factors could be involved. The aim of this research is to identify the architecture of the promoters and gene regulatory modules of the *dlg* gene. We will emphasize in the identification of regulatory modules that might explain the neuronal expression of the gene. A regulatory module is a region of DNA that binds multiple transcription factors, and in which their combined influence produces a specific expression pattern. The hypothesis is that these modules are in regions close to the transcription start. The aim of this work is to identify the modules (set of motifs), that characterize each of the patterns of expression. For this identification we performed a bioinformatic approach in multiple steps that starts with a comparison of the *dlg* genomic sequence of the twelve orthologs from the available genomes of the *Drosophilidae* family to select conserved motifs. The selected conserved motifs were compared to databases of known transcription factors binding sites. Having characterized the putative gene regulatory modules of *dlg*, we proceeded to search in the whole genome to identify genes that share a similar regulation, in particular we search for genes of neuronal expression. Finally using in-silico predictions, a vector with the putative modules upstream and the nuclear GFP sequence will be created in order to test in vivo the module capacity to drive expression in a specific tissue.

We found 4 modules independently tested experimentally. Two modules that would regulate DLGS97 and the others two DLGA. Modules 1 and 2 in transgenic embryos of *Drosophila melanogaster* would regulate the expression of DLGA and

modules 3 and 4 the DLGS97 expression. In transgenic *Drosophila melanogaster* larvae, modules 1 and 4 would regulate expression of DLGA. In adult transgenic modules 1, 2 and 4, would regulate the isoform DLGS97. The conclusions were obtained after dissecting each of the stages of development of *Drosophila melanogaster* and by observation in fluorescence microscopy.