

**Title:** Tuples, Tables, and Trees: A New Approach to the Discovery of Patterns in Biological Sequences

**Authors:** D.R. Argentar<sup>1</sup>, K.M. Bloch<sup>1</sup>, H.A. Holyst<sup>1</sup>, A.R. Moser<sup>1</sup>,  
W.T. Rogers<sup>1</sup>, D.J. Underwood<sup>2</sup>, A.G. Vaidyanathan<sup>1</sup>, J. VanStekelenborg<sup>1</sup>

**Affiliations:**

- 1) DuPont Company
- 2) DuPont Pharmaceuticals

**Email:** David.R.Argentar@usa.dupont.com, Karen.M.Bloch@usa.dupont.com,  
Herb.A.Holyst@usa.dupont.com, Allan.R.Moser@usa.dupont.com,  
Wade.T.Rogers@usa.dupont.com, Dennis.J.Underwood@dupontpharma.com,  
Ganesh.Vaidyanathan@usa.dupont.com, John.Van-Stekelenborg@usa.dupont.com

**Abstract:**

Genes and proteins are described by sequences of symbols representing their nucleotide or amino acid building blocks. Patterns of symbols in the sequences are linked to biological function, but are often difficult to discover, especially in cases where homology among the sequences is low. Nevertheless, finding functional loci via analysis of sequence data is key to every life sciences area, from the engineering of end-use properties in crops to the discovery of potent new drugs.

There are many problems in the analysis of biological sequences that are not adequately addressed either by alignment or matching methods. Most often, these problems seem to be characterized by low or undetectable homology at the sequence level, despite evidence of structural or functional similarity. This is the problem domain we are seeking to address.

In this poster we will describe a new pattern discovery algorithm based upon a novel set of data structures ("k-tuples", and associated tuple tables) combined with a new tree-traversal method, designed to efficiently discover all patterns at all levels of support.

Our algorithms are designed to address three major categories: pattern-based sequence clustering, motif discovery, and sequence/structure mapping. We will present results from the application of these methods to analysis of G-protein coupled receptors.