NEUROTECHNIQUES

Hunting hormones

Neuroscience Gateway (March 2007) | doi:10.1038/aba1724

A bioinformatics technique identifies two new peptide hormones in the human proteome.

Bioinformaticians have a message for peptide hormones: you can run, but you can't hide. Most peptide hormones contain fewer than 100 amino acids and act at G-protein-coupled receptors. Many act as both hormones in the periphery and neurotransmitters in the brain. Roughly 27 G-protein-coupled receptors lack known ligands, suggesting the existence of as-yet unidentified peptide hormones. Now Mirabeau *et al.* isolate new peptide hormones from a database of human protein sequences in a recent article in *Genome Research*.

Hidden Markov modeling assigns probabilities to states and their transitions, creating a model of the most likely path. For example, in the English language, the letter q (state) is followed with high probability by the letter u (transition). In bioinformatics, states are generally nucleotide or amino acid sequences or motifs. Because hidden Markov modeling assesses the probability of moving from one state to another, it scores sequences independent of their length.

The authors developed their hidden Markov model based on features common to secreted peptide hormones. Peptide hormones contain signal peptide sequences that are cleaved off before secretion. The authors trained the model on 1011 proteins that contain signal peptides from the SWISS-PROT database. Most peptide hormones lack transmembrane domains and contain amino acid sequences common to extracellular proteins. The authors trained the model to be similar to annotated sets of extracellular peptides and dissimilar from annotated sets of intracellular and transmembrane proteins. Mature peptide hormones are usually cleaved from longer pro-hormones. The authors trained the model on pro-hormone cleavage sites in the [MEROPS](#) peptidase database. They used a standard technique called the Viterbi algorithm to combine these features into a model describing the most probable state path (amino acid sequence).

How well did the model match actual peptide hormones? The model recovered known peptide hormones from databases containing roughly 30,000 protein sequences. The 300 highest scoring proteins obtained from the [Ensembl](#) database of human, mouse, rat and dog protein sequences contained 90% of the 75 known peptide hormone precursors. Of the top 300 proteins identified from the SWISS-PROT/TrEMBL human protein database, 26% were known peptide hormones and neuropeptides. To best identify new peptide hormones, the authors focused on 61 sequences in this best match list that were hypothetical or poorly annotated proteins.

The authors used several criteria to limit their list of 61 to the 2 most promising candidates, which they named spexin and augurin. When transfected into pancreatic cells, both spexin and augurin localized to secretory dense core granules. *In situ* hybridization localized spexin mRNA to the stomach and esophagus, and purified spexin induced dose-dependent contraction of stomach muscle, suggesting that it is involved in digestion. In contrast, *in situ* hybridization localized augurin to the intermediate lobe of the pituitary, adrenal, choroid plexus and heart, suggesting that it is involved in energy balance or cardiovascular function, according to the authors.

Additional investigation of the matches identified with this hidden Markov model and hidden Markov models designed to detect unique protein products from known pro-hormones should identify more candidate peptide hormones. Hidden Markov modeling should help identify new members of other protein families with defined sequence motifs, such as ion channels and G-protein-coupled receptors.

Debra Speert

1. Mirabeau, O. *et al.* Identification of novel peptide hormones in the human proteome by hidden Markov model screening. *Genome Research* (2007).