

BLAST into the Future of Biological Research



Deriving Function from Information

You've had a chance to explore the modeling tool Cn3D, now we're going to use what you've seen to try and understand the relationship between protein structure and function.

Given that the genetic material contained within DNA is responsible for creating the multitude of proteins that make up the intricate network of the human body, it may not come as such a surprise that certain protein structures and amino acid sequences are repeatedly utilized to perform particular functions in your body. In fact, this area of investigation is so crucial to our understanding of biological systems that it consumes a large portion of our government's scientific research expenditure. Today, you will be conducting your own experiment by comparing receptor proteins of two different hormones in an effort to see the similarities and differences between the structures of these proteins and to postulate functional differences which result from changes in amino acid sequence.

<http://www.ncbi.nlm.nih.gov> will serve as our research homepage.

Once at this site use the protein search tool to find the following proteins:

<i>Homo Sapiens</i> Protein	Search Criteria	Locus ID#
<i>ESR1</i>	<i>estrogen receptor 1, homo sapiens</i>	
<i>GR</i>	<i>glucocorticoid receptor, homo sapiens</i>	

From the records, note the length of each polypeptide (number of amino acids).

Now to see a multidimensional view of these polypeptides! Open the files on your desktop corresponding to the ESR1 and GR proteins and examine their structures side by side. The peptide is sequestering its respective ligand, indicated as a small molecule in stick mode, in the binding site. Manually determine the amino acid sequences surrounding the site. You can do this by highlighting the amino acid sequence in the "sequence toolbar" below the structure. The sequence colors correspond to structural domain colors.

Are the proteins similar in secondary structure (alpha helices, beta sheets)? How do they diverge? How different are the binding sites? Are they relatively similar in terms of size?

Now let's test your binding site hypotheses using the BLAST tool. On the bottom left of the BLAST menu, click the "align two sequences" option; this will allow you to compare the protein domains and give you a better idea of function.

Highlight the sequences surrounding the binding pocket from the sequence toolbar. Paste these protein sequences into each sequence box, then click on BLAST. You will obtain identity percentages when you align each sequence combination; carefully observe the strings of sequence that are conserved in each alignment. Repeat the BLAST using the protein IDs from above, to check the overall homology of the proteins.

Hormone binding site sequences are most often highly conserved in the protein families that contain them. Let's determine if your selected sites fit this assumption. From the BLAST menu, select "search for short nearly exact matches." Paste your predicted site sequence into the box provided and click BLAST. Your results will list proteins that are similar to your polypeptide sequence (ESR1 or GR binding site). These can include receptors for the same hormone, or for structurally similar ligands; judging by the list you receive, could this be the hormone binding site?

Protein	Binding Site Homology (ESR1-GR1)	Entire Protein Homology (ESR1-GR1)	Binding Site Homology (other proteins)
<i>ESR1</i>			
<i>GR</i>			

Researchers routinely perform searches for particular ligand binding pockets; however, the method we used takes far too long if one is to analyze several proteins! Instead, NCBI provides a summary of conserved binding domains for each protein of interest. To access those for ESR1 and GR, go to the protein record once more, and look at the top right-hand corner of the page. The three links provide more interesting information about your protein. Select "domains", and your protein will be delineated as a straight line (note the length), with attached domains. Click on a domain for a description of its hypothesized function. Does your predicted binding site fall into one or more conserved domains?

Protein	Domains	Location	Function
<i>ESR1</i>			
<i>GR</i>			