



# Master project in bioinformatics: protein domain versatility

at Stockholm Bioinformatics Centre, Albanova

**Background:** Proteins are composed of higher-level building blocks called domains, which often fold independently and which can be found in different combinations in many different proteins. Some domains practically always occur with the same neighbours in proteins, or are only found alone. Others domains are found in very varied protein contexts, even after compensating for how common they are overall. Such domains are called mobile, promiscuous or versatile in various studies. There have been previous reports about which innate properties make a domain versatile, that is, to co-occur with many different types of domains, but their results are in disagreement.

**Aims:** The aim of this project is a comprehensive study of the possible functional bias of versatile protein domains, using the most recent data and within a stringent statistical framework. Using all complete proteomes present in the most recent version of the Pfam domain database, the versatility of each domain or clan should be computed within each proteome. The domains can be ranked within each proteome in order of versatility; this rank should be compared between species as well as between kingdoms to determine how consistent this property is. Furthermore, we want to investigate any potential functional bias of high-versatility versus low-versatility domains. For this purpose, functional associations, such as Gene Ontology annotations, of domains should be determined, and an enrichment analysis between the sets of domains should be performed, using suitable correction procedures to compensate for the vast number of hypotheses being tested. If this comparison yields relevant results, the project should result in a publication in a peer-reviewed journal.

**Requirements:** The applicant should be familiar with basic statistics (specifically significance testing of enrichment and multiple testing correction), basic bioinformatics scripting (we prefer Perl but accept other languages) and sufficient biology/bioinformatics to understand the question we want to answer. We work in a Linux/UNIX environment. The applicant should be able to work independently (but with supervision available on request) and must be prepared to verify, check and double-check their work as we must be completely sure the results are genuine if we are to be able to publish them. Please apply by sending your CV and the email address of a reference person to:

Erik Sonnhammer, Ph.D.  
Professor of Bioinformatics  
Head of Stockholm Bioinformatics Centre  
AlbaNova University Centre, Stockholm University  
S-106 91 Stockholm, Sweden  
Tel: +46-(0)8-5537 8567  
Email: [Erik.Sonnhammer@sbcsu.se](mailto:Erik.Sonnhammer@sbcsu.se)  
<http://sonnhammer.sbc.su.se/>