

Powerful Tool for Protein Analysis Provided Pro Bono

Contributed by IV Post
Thursday, 04 December 2008

Stanford, California - Scientists around the world may benefit from a powerful new database, available for free online, that will help them to home in on the parts of proteins most necessary for their function.

Arend Sidow, PhD, associate professor of pathology and of genetics at the Stanford University School of Medicine, recently launched the novel bioinformatics tool, which enlists evolution as the guide to determining the role different proteins play in a wide array of organisms.

ProPhyLER, which Sidow has been working on since 2002, will enable a researcher studying a protein, or the gene coding for it, to more easily figure out how it works and whether something might go wrong if the gene has a mutation. "Whether you're a cell biologist, biochemist or structural biologist, ProPhyLER produces instant working hypotheses for you as to where the protein's functional areas are," he said. The site made its debut on October 10.

Proteins - the machines of life that do everything from making your muscles move to helping you breathe and think - are long chains of small chemical units called amino acids. As soon as a protein molecule is made inside a cell from the gene encoding it, it folds up to assume the unique shape that determines its activity. To do its job, a protein needs to have specific amino acids (there are 20 to pick from) in specific places. In particular regions of the folded protein, it may be crucial that a specific amino acid sequence be there in order for the protein to function; in other regions, the swapping of one amino acid for another has little effect.

Over the course of several hundreds of millions of years myriad species have evolved, and, through eons of random mutation, so have their proteins. Yet in the face of all these changes, some things have to remain constant. "Evolution imposes stronger constraints on more-important regions of a protein molecule, from the standpoint of its biological activity, than on other, less-critical regions of that protein," Sidow said. If a change interferes with a protein's function, the hapless creature harboring this variant dies out; if not, the creature is fruitful and multiplies, and the variant protein persists in modern species.

It's by no means obvious just from looking at a protein's linear amino-acid sequence which regions are the "business districts" of the protein, and which are the sleepier bedroom communities. However, ProPhyLER shows biologists which parts of a protein are key to its activity by comparing numerous versions of the same protein from different species. This is especially useful for proteins about which nothing or little is known, which is still the majority of proteins encoded in the human genome.

Human geneticists, too, will benefit from ProPhyLER (a play on words derived from the bulkier term "PROtein PHYLogenetics and Evolutionary Rates"). Each of us carries tens of thousands of protein variants (due to mutations that have persisted in the human gene pool), some fraction of which affect protein function. For researchers, it is notoriously difficult to measure experimentally how much a protein's function is impaired by a mutation. ProPhyLER provides specific predictions, also based on evolutionary variation, of the impact of a mutation on the protein's function. A mutation in an amino acid that has changed a lot in evolution is much less likely to be bad than a mutation that affects an amino acid in which evolution has not tolerated any change. "This type of analysis will be key in the interpretation of your personal genome sequence, when - not if - that becomes commonplace," said Sidow.

After a user has searched for his or her protein of interest, the ProPhyLER Web site displays data via two interfaces. The first displays all the evolutionary data graphically along the length of the protein. The second, called "Crystal Painter," projects these degrees of evolutionary constraint onto three-dimensional structures of proteins, when those are known, by imposing a color-coded scheme on their structures. It is clear at first glance that parts of proteins obviously important to function - such as binding pockets in which the protein holds a small molecule on which it's performing an operation - are, indeed, just the ones Sidow's evolutionary algorithm has deemed most important.

“No other proteomics resource does this,” Sidow said.

The National Institutes of Health/National Human Genome Research Institute provided funding for this project.