Master’s Thesis Abstract

UNIVERSITY OF TURKU
Department of Information Technology

Tiwari, Avlokita: Phylogenetic Analysis of Avidins and Expression of Novel Avidins from Lactrodectus hesperus and Hoeflea phototrophica
Master’s thesis, 64 p., 10 appendix pages
Master’s Degree Programme in Bioinformatics
October 2015

Avidins (Avds) are homotetrameric or homodimeric glycoproteins with typically less than 130 amino acid residues per monomer. They form a highly stable, non-covalent complex with biotin (vitamin H) with $K_d = 10^{-15}$ M (for chicken Avd). The best-studied Avds are the chicken Avd from Gallus gallus and streptavidin from Streptomyces avidinii, although other Avd studies have also included Avds from various origins, e.g., from frogs, fishes, mushrooms and from many different bacteria. Several engineered Avds have been reported as well, e.g., dual-chain Avds (dcAvds) and single-chain Avds (scAvds), circular permutants with up to four simultaneously modifiable ligand-binding sites. These engineered Avds along with the many native Avds have potential to be used in various nanobiotechnological applications.

In this study, we made a structure-based alignment representing all currently available sequences of Avds and studied the evolutionary relationship of Avds using phylogenetic analysis. First, we created an initial multiple sequence alignment of Avds using 42 closely related sequences, guided by the known Avd crystal structures. Next, we searched for non-redundant Avd sequences from various online databases, including National Centre for Biotechnology Information and the Universal Protein Resource; the identified sequences were added to the initial alignment to expand it to a final alignment of 242 Avd sequences. The MEGA software package was used to create distance matrices and a phylogenetic tree. Bootstrap reproducibility of the tree was poor at multiple nodes and may reflect on several possible issues with the data: the sequence length compared is relatively short and, whereas some positions are highly conserved and functional, others can vary without impinging on the structure or the function, so there are few informative sites; it may be that periods of rapid duplication have led to paralogs and that the differences among them are within the error limit of the data; and there may be other yet unknown reasons. Principle component analysis applied to alternative distance data did segregate the major groups, and success is likely due to the multivariate consideration of all the information. Furthermore, based on our extensive alignment and phylogenetic analysis, we expressed two novel Avds, lacavidin from Lactrodectus Hesperus, a western black widow spider, and hoefavidin from Hoeflea phototrophica, an aerobic marine bacterium, the ultimate aim being to determine their X-ray structures. These Avds were selected because of their unique sequences: lacavidin has an N-terminal Avd-like domain but a long C-terminal overhang, whereas hoefavidin was thought to be a dimeric Avd. Both these Avds could be used as novel scaffolds in biotechnological applications.

Keywords: Avidins, phylogenetic tree, multiple sequence alignment, MEGA, protein expression, principle component analysis