verification of the quality of the CD data deposited, in order to maintain a high standard of data incorporated into this resource. The PCDDB will accommodate both conventional (lab-based) CD data and synchrotron radiation circular dichroism (SRCD) data. It will provide for easy wide-spread dissemination of research results and a facility for data sharing, as a simple means of fulfilling granting body requirements. The first release of the Protein Circular Dichroism Data Bank was made publicly-available in September 2009. The first entries were of the 71 proteins that comprise the SP175 reference database (Lees et al., Bioinformatics 22:1955; 2006). In the current release, users can download and access these data files. In the full beta release, due in late 2009/early 2010, users will be able to deposit their own spectra, which will be validated before being made openly available on line. The PCDDB can be accessed through the website http://pcddb.cryst.bbk.ac.uk.

(Supported by a grant to BAW and RWJ from the BBSRC Bioinformatics and Biological Resources Fund).

1022-Pos

Entropic Fragment Based Approach for Aptamer Design

Chih-Yuan Tseng, Jack Tuszynski.
University of Alberta, Edmonton, AB, Canada.

Aptamer, a short RNA/DNA sequence, is designed through SELEX (systematic evolution of ligands by exponential enrichment) to bind to specific targets including small molecules, proteins, nucleic acids, and even cells, tissues and organisms. Several advantages such as binding specificity and affinity and non-toxic and non-immunogenic properties make aptamer a promising tool in therapeutic applications. Basically, SELEX starts with preparing a pool of random RNA/DNA sequences and consists of a series of enrichment processes. In each step, the process will identify sequences that have the highest binding affinity. The success of SELEX hinges on synthesizing ''good'' random sequence pools. A ''good'' pool should have sufficient sequence diversity and structural complexity. Furthermore, the quality of sequence pools also greatly influences efficiency of SELEX. These criteria discourage the application of the conventional virtual screening approach.

Therefore, we propose an entropic fragment based approach that is free from these criteria to design aptamers given a target protein in this work. The crux is to introduce probabilistic description. First, the approach utilizes limited information such as the interactions of nucleotide fragments and target proteins to determine the probability of having such interactions. Afterward, based on the method of maximum entropy (ME), the preferred nucleotide fragment that mostly likely interacts with target proteins is the one that maximizes the entropy of the system. By repeating the same procedure given the fragment determined in previous step, a preferred aptamer then can be constructed. At last, we consider the thrombin aptamer designed from SELEX as a target to investigate the applicability of the proposed approach.

1023-Pos

Transcription Factor-Target Gene Mapping Enhanced by Integrating Motif Search, Function Annotation and Expression Data

Yu Bai.
Seralogix LLC, Austin, TX, USA.

Transcriptional regulation is essential for all eukaryotes. Defining regulatory networks, linking transcription factors (TF) to target genes, is of fundamental importance to biology. Developing predictive models for TF-target mapping is critical to test current understandings and to propose new hypotheses. Nevertheless, predicting target genes remains challenging because TF binding motifs are often short, degenerated and widely-spread. Moreover, the binding mechanisms are generally more complex than recognizing a specific sequence. Thus, apart from leveraging the motif identification accuracy, it may be helpful to integrate alternative knowledge depicting the relationship between the TF and the targets.

Herein we developed an integrated TF-target mapping strategy and examined its performance. The candidate genes were predicted by a sum of three factors: the enrichment and quality of the putative motifs, identified via an optimized position weighted matrix (PWM) score over phylogenetically conserved promoter regions; a Gene Ontology-based semantic functional relevancy measure; and a causal relationship measure between the gene and the TF derived from expression profiles. The evaluation was conducted using 52 transcription factors covering most of the known, PWM-available TFs in higher eukaryotes, and their total of 1315 curated target genes.

The integrated strategy achieved a considerably higher accuracy with an area under receiver operating characteristic curve (AUC) of 0.67, compared to the commonly-adopted method relying on solely a motif enrichment score (AUC of 0.56). In particular, optimizing the PWM score in phylogenetically conserved promoters increased both sensitivity and specificity; functional relevancy and causal correlation further lowered the false positives.

The results illustrate the feasibility of integrating multiple knowledge sources to improve TF-target mapping. The presented strategy is readily scalable to genome-wide, and can be applied along with other inference tools to assist the regulatory network reengineering.

1024-Pos

Prediction of Functional WXXF-Like Protein Motif from Sequence

D. S. Dalafave.
The College of New Jersey, Ewing, NJ, USA.

Development of reliable techniques to predict functional peptide motifs from their sequences is an important task in bioinformatics. Identification of functional motifs through experiments can be difficult, while searching protein databases often requires substantial computational resources. Finding short motifs is particularly hard. Presented here is a simple and effective way to identify short, functional WXX(F/W) motif from sequence. The method involves a development of a scoring function based on the sequence properties. Clathrin-coated vesicles coordinate selective transport of molecules across the membrane. The AP-2 adaptor complex is essential for the clathrin-coated vesicle formation. Two ''ear'' domains of the AP-2 complex bind to clathrin and accessory proteins. The accessory proteins interact with the AP-2 via short motifs. One such motif is WXX(F/W). Substitutions of W in the first or F/W in the last position of the motif eliminate the binding with the ear domains. Residues in position 2 of the motif can be mutated to some extent, while position 3 is very flexible to residue substitutions. In this work, three-dimensional computer models of known WXX(F/W) variants bound to the ear domains were constructed. Systematic residue mutations were done to determine sequence properties crucial for the motif's interactions, and thus for its function. The sequence properties were used to construct a scoring function. The function was tested on randomly generated and other known WXX(F/W) sequences. The scoring function successfully captured relationships between sequences' properties and their functionalities. Several false positive results were obtained. However, the scoring function reliably identified nonfunctional sequences. New putative functional motif variants were predicted. This study on functional WXX(F/W) motif should increase our understanding of vesicle transport mechanisms.

1025-Pos

Functional Characterization of Tubby Domains of Arabidopsis Thaliana Using Computational Methods

Shaneen M. Singh, Nataraj V. Dongre.
Brooklyn College, Brooklyn, NY, USA.

Tubby domain containing proteins are cell signaling proteins common to many multicellular eukaryotes. The tubby domains have dual binding function and are capable of interacting with both DNA and phosphotidylinositol, thereby potentially functioning as transcription factors as well as membrane associated signaling factors. Structurally they adopt a barrel structure with a putative DNA binding groove that terminates in a phosphoinositide binding basic pocket as revealed in experimentally solved tubby domains. Tubby and tubby-like proteins have been implicated in the maintenance and function of neuronal cells during post differentiation and development, and mature-onset obesity in animals but not much is known about the function of tubby domain proteins in plants. We have undertaken a comprehensive computational examination of all tubby domains in the model plant, Arabidopsis thaliana to understand the structural basis for the mechanism of their function and to compare them with tubby-domains in other organisms. We have modeled the various tubby-domain proteins in A. thaliana, which share 30- 80% sequence similarities across their C-terminal tubby domains and are unique in possessing a conserved N-terminal F-box domain of about 50 residues in almost all members, using an automated modeling pipeline coupled with manual refinement methods. The biophysical analysis of all tubby domains in this model plant represents the first structural investigation of these domains in plants and provides initial insight and predictions as to their function in plants.

1026-Pos

HAMDAM-1 as a Sequence-Based Software for Studying the Physical Properties of Proteins

Hamid Hadi Alijanvand, Maryam Rouhani, Ali A. Moosavi-Movahedi.
Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran, Islamic Republic of.

In the field of protein evolution there are a few studies that focus on the physical parameters of the protein. For a comprehensive study of physical parameters it is necessary to consider biothermodynamic parameters, structural and statistical properties simultaneously. We developed the HAMDAM-1 as a sequence-based software which is capable of calculating different physical parameters of proteins synchronously based on their amino acid sequences. Our results could confirm the co-evolution among three interacting proteins Cav 1, α-actinin and rSK2. The difference between rSK1 channel and the other proteins of this family (rSK2 and rSK3) were also authenticated.