Boojala V. B. Reddy

Statement of Teaching

I believe in teaching students not merely a subject. I work towards the development of each and every individual student and provide the best learning environment and to impart knowledge to students to the best of their abilities. Teaching is a value-laden activity that benefits both the students and the teacher. My teaching interests stem largely from my experience of many wonderful teachers throughout my graduate career. My desire to pursue a career in life sciences was in large part to the impact of my Biochemistry teacher who presented the concepts in a creative, simple and clear manner. The enormity of that impact a teacher can have in sharing his knowledge serves to motivate me to do the same.

I have found it to be a rewarding and learning experience teaching graduate students. It gave me an opportunity to carefully research and organize the material in order to present the theoretical concepts behind various protein sequence data analysis tools. Since 1989 I have enjoyed my teaching in in CCMB and UCSD which has also helped more in my research activity. Teaching the subject served also to expand my research horizon. Thus, I strongly believe in the integration of research techniques and methodologies in the classroom.

Teaching students in this multi-disciplinary area of bioinformatics, it is necessary to have a comprehensive knowledge of the subject as well as related areas (e.g., molecular biology, biochemistry, information systems, computational science, molecular modeling and simulation). My strong background in protein structure analysis, modeling, informatics and related scientific disciplines allows me to plan my lectures and integrate essential elements of these related fields. I emphasize the understanding of basic concepts in bioinformatics, the science behind the development of tool, and application of their theoretical knowledge to biotechnology and drug industry. I endeavor to make the curriculum interesting and exploratory through projects utilizing a variety of tools in solving interesting bioinformatics problems.

As an instructor at the UCSD bioscience department, I have developed course curriculum to teach “Protein Sequence and Structure Data Analysis Tools and Algorithms”. In this course students are provided with an opportunity to explore a number of different bioinformatics tools and apply it to practical problems in the field. I have developed a special course on “Principles of Protein Structure and Comparative Modeling” aimed at the industry professionals attending UCSD. At U of MN I have now designed a course “Perl for Bioinformatics” to be taught in the summer Bioinformatics workshop.

Students have diverse backgrounds and learning styles depending on their background (educational and cultural) I pay utmost importance to understanding and tailoring the material to make it suitable for students so they may learn in a style that is most comfortable for them. In addition to individual learning, I encourage students to work together via group study and projects. I have found it advantageous to utilize a method of continuous assessment throughout the course. This can be done through homework assignments, quizzes, and short tests. This provides students the opportunity to quickly assess their learning skills and build on the course material step by step.

My background and experience with students has given me a keen ability to listen carefully and respond to their needs in a positive manner. I endeavor to maintain an open door policy and be easily accessible to my students via personal contact, phone, and email. A course website is a useful resource for communicating the information as well as an assessment tool. I strongly encourage students to utilize resources available through modern technology including the internet. Classroom lectures may also be recorded and provided on the course website to assist students to review the classroom instructions.

I like to facilitate a free and open dialog with the students so they feel comfortable to provide me with their comments as I share mine with them. Students may send in their comments via the course website or directly via email or a personal note.

----------oOo----------

Statement of Research Interests

Over the past 15 years I have been working in the area of computational analysis of DNA and protein sequence and structure data, developing computational tools and web servers. With the recent initiatives of genomic sequencing, high throughput structure determination and microarray techniques have been generating wealth of biological database. The key issues in building reliable knowledge-bases (expert systems) are efficient data mining and finding the right differentiators that are unambiguous, quantifiable and sufficiently robust to deal with the heterogeneous datasets common to biological systems. I consider the field of bioinformatics is very challenging since it involves a combination of tasks such as thorough understanding of biological processes, expert-level data mining, designing the algorithm to extract information, develop predictive methods and making the resultant tools and databases accessible to the public. I developed several useful computational methods using both sequence and structure data analyses and have several well cited research findings as summarized below.

BACKGROUND AND BIOINFORMATICS EXPERIENCES

My research publications covers broadly the DNA, protein sequence and structural data analyses, protein modeling, structure determination and docking studies and falls in the area of "Computational Biology and Bioinformatics". (please see list of my publications for the references cited below).

A. Protein Structure Data Analyses:

Protein-Protein Interactions: I am presently working on protein-protein interaction studies. I have analyzed several known protein-protein complex structures and computed various parameters that help in predicting the interaction sites between the two protein structures. I have done docking calculations, computed pair potentials, desolvation energy, conservation index and several other parameters to develop computational method to predict interacting surface regions of protein structures (Reddy *et al.* 2004; Duan *et al.*, 2004).

Common Substructures: My work on analysis of common substructures in proteins gives a method that utilizes structure based sequence alignments in the protein data bank (PDB) and protein sequence data base (SWALL, NR). The method identifies conserved key amino acid positions (CKAAPs) for each commonly occurring substructures and representative protein structures in the PDB (Reddy *et al.*,2002*;* Reddy *et al.*, 2001). Using the method we have developed a CKAAPs resource data bases (Li *et al.*, 2001, Li *et al.*,2002). CKAAPs are proposed to be useful in fold recognition and protein engineering.

Secondary Structural Packing: I have analyzed packing of secondary structural elements in proteins. The analysis was done on helix-helix, helix-sheet, and sheet-sheet packing in protein structures. The results were shown to be useful to improve comparative protein modeling procedures. The work helps in adjusting rigid secondary structural elements, helices and sheets in the core of the protein models (Reddy, (2002; Reddy &Blundell, 1993; Reddy *et al.*, 1999; Nagarajaram *et. al*, 1999).

Protein Stability: We have evaluated the structural environment of amino acids in known protein structures. About 360 substituted single residue mutants were collected from the literature for which parent protein structure and altered *in vitro* (thermal, solvent or pH induced) stability information was available. The structural environment of each of the parent residue is characterized and a method is developed to suggest substitution mutations to engineer *in vitro* stability of a given protein sequence for which structure is available (Reddy *et al.*, 1998).

Modeling Studies: I have done comparative modeling of structure of rabbit M4-Lactate Dehydrogenase based on the available structure of dogfish LDH-M4 (Rajenderkumar *et al.*, 1994). I participated in the CASP3 test as member of Tom Blundell's group (Burk *et al.*, 1999) and succeeded with best models. I also submitted CASP3 targets for secondary structure prediction using the method we developed at CCMB (Tiwari & Reddy, 1999). The modeling procedure for DNA-protein interaction is developed and carried out using homology based prediction and chemical and stereo-chemical rules from the knowledge of known X-ray/NMR structures of DNA-protein complexes. A general method of approach is

standardized to model sequence dependent DNA binding. Using this approach we modeled a DNA protein complex, -35 hexamer of promoter interaction with 4.2 helix-turn-helix domain of sigma-70 subunit of *E. coli* polymerase and its mutants V576G, V576T (Reddy *et al.*, 1997).

Peptide Modeling and Drug Design: I have done comparative modeling of small peptides that were identified to mimic scatter factor and FGF. Later small molecular compounds were designed that are analogous to the peptide model structures. At least four of the compounds were tested experimentally and found to be potential drug compounds by a New York based company.

B. Protein Sequence Data Analysis and *in vivo* Protein Stability:

Proteins are known to degrade rapidly when conformations are altered due to abnormality in the sequences. Normal cellular proteins also display a wide range of half-life - turnover rates of individual proteins can differ as much as 1000-fold. Sequence specific properties, global features and the location of a protein in the cell are found to be important in deciding the intracellular stability of a protein. In order to identify sequence dependent propertied we have analyzed the stable and less stable protein sequences and observed that the di-peptide composition in stable proteins is significantly different from the less stable proteins This observation was used to develop a theoretical method to predict protein stability from its amino acid sequence information (Guruprasad *et al.,* 1990). We further analyzed the structural location of these di-peptides in non-homologous protein structures and found that the general distribution of sensitive (stable and less stable) di-peptides is high in regions that are solvent accessible and have more hydrogen bonding interactions with near neighbor residues. These di-peptides are also usually present closer to the molecular surface and significantly more occurrence in turns and random coil structures (Reddy, 1996). Based on these structural observations we developed a method that gives theoretical suggestions for substitution mutations to alter the intracellular stability of a given protein. Intelligenetics USA (PHY_CHEM) has integrated our algorithm into the sequence analysis software of PC-GENE package. The method further refined and developed as bioinformatics tool to suggest substitution mutations to increase intracellular stability of proteins (MEICPS: Reddy *et al.,* 1999).

C. DNA Sequence Data Analyses:

(i) My research career started with analyses of DNA sequence data to understand synonymous codon usage (Kolaskar & Reddy, 1985a). I have developed a method to locate protein coding sequences in DNA of prokaryotic systems (Kolaskar & Reddy,1985b). This algorithm has been integrated and used as part of DNA sequence analysis software of PC-GENE package by Intelligenitics USA (COD_PROC). I have further studied the contextual constraints on codon pair usage and also synonymous codon usage in the DNA sequence (Kolaskar & Reddy, 1886; Kolaskar *et al.,* 1995).

(ii) I have also analyzed the eukaryotic DNA sequences of Mice and Human genes. I developed a statistical analytical method to predict splice-sites in these DNA sequences. The method has given best prediction results with much less false positives compared to the other methods available during the time (Reddy *et al.*, 1991; Reddy & Pandit, 1995).

FUTURE RESEARCH INTERESTS

The advances in genome sequencing, high throughput structure determination and microarray technology have been generating vast amount of data and its utility is smoothened by the information technology to make it readily available in different forms as knowledge-base to the scientific community. Nearly 600 genomes are sequenced and over 25,000 protein structures are available and several depositions of microarray expression data for genes of developmental and deceased states are available in the public databases. With the presently available computational and molecular biology tools it is possible to track the proteins from the expression to a structural model and to the design of therapeutics. In a broader sense I am interested to initiate several collaborative projects with pure molecular biologists to use my expertise and utilize the available computational tools, knowledge-base for therapeutic process and for understanding the basic science in terms of molecular functions and interactions.

I am further interested in continuing my bioinformatics research activity for the genome-wide analyses. I plan to expand my research activities and develop research projects for microarray data analysis and use

comparative genomics techniques to mine the information in the gene expression data. I have few specific research projects in mind where progress has been minimal such as:

Genome-wide proteomics studies – Yeast two-hybrid assays and high-throughput mass spectrometry – provide a growing list of putative protein-protein interactions. Many biologically important interactions occur in weak transient complexes that are not amenable to experimental analysis. I have been interested to work on bioinformatics of protein-protein interactions to explore possible ways if identifying interacting partners from the sequences and structure information.

Hyperthermophilic, thermophilic and mesophilic bacteria and archaea reveal subtle sequence and structure-dependent features that are responsible for their varied stabilities which I propose to exploit. To analyze all available homologous protein structures from hyperthermophilic, thermophilic and mesophilic proteins. I compute amino acid residue structural environments in each of the proteins and compare the variations. This analysis will help us to identify structure-based reasons for the varied stability of proteins. I would further analyze the sequence data of proteins with respect to the thermal stabilities at which they function with optimal efficiency. We will study the sequence variations that define the stability of protein structure within the homologous variations. Finally, I propose to use all identified sequence and/or structure-based features of protein stability to develop computational methods to suggest substituted mutations that engineer protein stability.

Comparative protein modeling (CPM) is ability to model protein structures for the sequences using a homologous structure from structural data. CPM is providing new breakthroughs in our ability to use the large amount of sequence data coming from the genome projects. Protein structures modeled through CPM improves as the number of experimental template structures increases, an event accelerated by the recent structural genomics initiatives. However, further improvements are necessary in various steps of CPM to obtain more useful and accurate models. The current state of the art fails to provide a model of the target that is closer to the real structure than the template it is modeled against. I propose to develop better tools for the protein core modeling step of CPM. It will do so by more appropriately arranging the packing orientations of rigid secondary structural elements to better define the target model thereby reducing the bias towards the reference template(s) used in modeling. More accurate protein models are useful to elucidate protein sequence-structure-function relationships, and for drug design experiments.

Recent structural genomics initiatives are exponentially increasing protein structural data. The next challenging task for structural bioinformatics is to predict interacting sites which are useful in modeling the interactions and identifying the drug target sites. I am interested to initiate research in this line of thought through analysis of known protein complex structures and protein-protein interaction data to model protein structures and their interaction sights and collaborate with experimental biologists for drug-designing studies.

I am an expert in small molecular design, screening and search tools, comparative protein modeling and docking studies. I therefore have potential to generate collaborative projects with pharmaceutical and biotechnology companies for drug design studies. I have worked on angiogenesis inhibitor design and c-met activator and inhibitor design studies and peptide mimetic design for a couple of local pharmaceutical companies.

In summary, (1) I plane to develop computational methods for the analysis of biological data bases for high-throughput analysis. (2) Develop customized algorithms to address specific scientific problems through data mining. (3) Develop and improve methods to model protein structures and interaction for therapeutic studies and for understanding the functions. (4) Establish collaborations and develop internationally visible and externally funded research programs in the above mentioned research areas. (5) Develop teaching curriculum in bioinformatics and computational biology. (6) Design suitable research projects for graduate, undergraduate students.

--------oOo--------