

Chemoinformatics

ADCOM 2009 Tutorial

17/12/2009

Forenoon

Lecture 1 of 3

Chemoinformatics: general concepts and applications

The aim of chemoinformatics is to gather new chemical information exploiting existing knowledgebase, and computational techniques. Computers, however, do not understand chemical compounds the way we do, so a major challenge in chemoinformatics is about chemical representation. This problem has been solved using various approaches, such as developing 1D, 2D, 3D molecular descriptors, or representing molecules as graphs, or line notations. Once a chemical representation of a compound has been achieved in a computer, various questions can be asked and analyzed. For example, how similar are two (or more) compounds based on structure, topology, reactivity or other physico-chemical properties? To answer this question a variety of 1D, 2D and 3D matching algorithms have been developed. Similar algorithms have been adapted for searching databases and data sources in chemistry.

Chemical compounds are not static entities; they undergo conformational transitions that modulate their property. Understanding the structural dynamics and correlating it with the consequent change in the chemical property is an important area of study in chemoinformatics.

Chemoinformatic studies also focus on analysis and prediction of interactions between small molecules (also self interaction) and with biopolymers. A very popular application of such studies is in virtual drug discovery.

In this talk I will introduce broad concepts on Chemoinformatics and illustrate some of them through example applications.

Target audience. The tutorial is suitable for students in any discipline with background in chemistry and exposure to computers.



Dr. Debnath Pal, is an Assistant Professor at the Bioinformatics Centre, Indian Institute of Science, Bangalore, India. His research interests include analysis of structure-function relationships in proteins and nucleic acids, mass spectrometry, and application development in computational biology. He has published over thirty research papers in major international journals in the area of chemistry and computational biology. He has received fellowship from the Alexander von Humboldt Foundation, Germany, and the Young Scientist Medal from the Indian National Science Academy, New Delhi.

Chemoinformatics

ADCOM 2009 Tutorial

17/12/2009

Forenoon

Lecture 2 of 3

Chemoinformatics in the post-genomic era

Recent advances in cancer and meta-genome sequencing studies have generated a wealth of sequence data, which when fully analyzed can potentially transform our understanding and treatment of human diseases. However, the computational challenges associated with integrating and analyzing large sequence datasets has posed major challenges in chemoinformatics, which seeks to translate genomic discoveries into therapeutic strategies. In this talk, I will describe some of the computational tools and approaches we have developed to extract biologically meaningful patterns from large sequence datasets. I will use the protein kinase gene family to illustrate how combining sequence data with structure and functional data can provide new insights into proteins underlying mechanisms.

Target audience. The tutorial is designed for students, post-docs and computational scientists interested in using bioinformatics and chemoinformatics approaches to understand protein structure and function. In particular those interested in large and functionally diverse protein families are encouraged to attend.



Dr. Kannan is an Assistant Professor in the Department of Biochemistry and Molecular Biology, and Institute of Bioinformatics at the University of Georgia, Athens, USA. He has published over 25 papers in major international journals in the field of bioinformatics and comparative genomics. He is the recipient of the Georgia Coalition Distinguished Cancer Scholar Award and the Indian National Science Academy Young scientist award. His research at the University of Georgia is focused on understanding the underlying mechanisms of biomedically important proteins, such as protein kinases and phosphatases, using comparative genomics and systems biology techniques

Lecture 3 of 3

Computational Protein Design

The ability to precisely engineer enzymes or hormones that are stable under varied conditions or even create new ones that can perform entirely novel functionalities has great scientific value. In earlier days of de novo protein design, the complex task of assigning a sequence to a target structure has been achieved through rational attempts. The rationale was the collective knowledge of differential conformational behavior of amino acids in a protein structure, accumulated from protein folding studies. Although plenty of efforts have already been taken towards the rational protein design, the success rate has not been very encouraging. Partly it is due to the lack of specific knowledge in the sequence coupled selection of protein backbone conformation.

In computational protein design, energy functions and combinatorial search methods has been employed in optimizing the sequence of a polypeptide fold. The automated design strategy is intended to minimize the combinatorial-search load in sequence selection and thus involves discretization and coarse graining the conformational-search space of side chains. The broad consensus in the field is to use Lennard-Jones energy, hydrophobic reward and desolvation penalty, Coulomb energy, an entropy term scaling the side-chain usage in the native-like frequency and orientation-dependent and hybridization-specific hydrogen bond energy as components of energy function. Variants of Dead End Elimination (DEE) and Monte Carlo search methods are used either independently or combined to optimize the sequence space.

Target audience: The tutorial will be helpful for researchers and academicians who are interested in making use of computational methods for addressing problems in chemistry and life sciences. In particular, it would benefit scientists engaged in research that involves molecular modeling, bio-molecular simulations, drug design, peptide and protein design etc.



Dr. Vibin Ramakrishnan is a Faculty Scientist in Institute of Bioinformatics and Applied Biotechnology (IBAB) Bangalore. He received his Ph.D from IIT Bombay in 2005 on the topic *Stereo chemical effects in protein structure, folding and de novo design*. He was the recipient of Innovative Young Biotechnologist Award 2007, given by Department of Biotechnology, Govt. of India. His research interests include characterizing biophore with reduced one dimensional descriptors and modeling folding pathways of complex proteins.