Message

I am happy to note that the first issue of “Recent Patents in Bioinformatics for Drug Research” is ready for online publication, a collaborative effort of BTIS Centre, Department of Biotechnology and Central Drug Research Institute, Lucknow.

“Bioinformatics” is an emerging field of science with wide applications in drug research. The process of designing a new drug using bioinformatics tools, opened a new area of research, growing very fast and adding new possibilities in drug discovery process.

Not all research gets published in papers and a lot of information is made available to the public through patents. A detailed analysis of patents granted in a particular area of research can provide valuable and missing information links. By analyzing the patents and studying the prior art, the research gaps can be identified, and the research work to be taken up can be focused. Thus a regular updates of patents is necessary for research, planning & development.

I hope through this bimonthly online periodical we shall be able to keep our researchers informed about the recent applications of bioinformatics in drug research.

Dr T. K. Chakraborty
Director
CDRI, Lucknow
Message

Bioinformatics has come a long way since it caught the attention of scientific community. The Application on Bioinformatics is widespread. Bioinformatics approach in modern biology has achieved remarkable success. This is quite evident in the ever-growing number of patents being filed and accepted on application of Bioinformatics in Drug Development.

I am happy to note that the CDRI Bioinformatics Centre as part of the Biotechnology Information System Network (BTISnet) is initiating a bimonthly online periodical aimed at providing the latest information on the patents evolved from all over the globe in the area of Drug Research. CDRI-BTIS Centre has always been active in disseminating information on Bioinformatics in the area of Drug Research through their other two publications. I am sure their new venture “Recent Patents on Bioinformatics in Drug Research” will provide great support for the scientific community especially for the Biomedical Researchers across the country.

Dr Madhan Mohan
Adviser
BTISNet
Department of Biotechnology
From Coordinators Desk………. 

The CDRI-BTIS Centre is happy to bring before you the inaugural issue of “Recent Patents on Bioinformatics in Drug Research,” a bimonthly online periodical. The periodical aims to provide a comprehensive coverage of Bioinformatics Patents related to drug research.

Bioinformatics is generally defined as the research, development & application of computational tools and approaches for expanding the use of biological, medical and health data, including those to acquire, store, organize, archive, analyze and visualize such data. This field of science came to the public attention as a result of the Human Genome Project, which generated a vast amount gene sequence data including human genome sequence data.

After the availability of human genome information, a new paradigm for drug discovery has emerged. The identification of the disease associated genes is the first step in drug discovery process, followed by determination of expression of genes in a given disease, understanding the function of gene products and finally to the development of drug candidate. In silico processes and tools between the various points of drug discovery are the focus of bioinformatics.

The promise of bioinformatics is that drug discovery can be accomplished in silico, that is, the computational analysis of the massive volume of biological data can identify the cause of disease and enable the design of sophisticated drugs to treat them. Bioinformatics has the potential not only to reduce the time and money required to develop drugs, but also to reduce the rate of failure of drug candidates and provide drugs that treat diseases more effectively and precisely.

Bioinformatics is one of the fastest expanding fields in Indian biotechnology sector today. There are over 200 companies in Bangalore, Hyderabad, Pune, Chennai and Delhi that are in some way involved in bioinformatics. Large scale IT organizations, both global and Indian, are also getting into this sector. Indian Government is extending full financial support to this industrial sector. India was the first country in the world to establish a nation wide bioinformatics network (BTIS network) under Department of Biotechnology (DBT).

“Patents” is a source of information that cannot be ignored in drug research and development. Patent protection is generally not sought for innovations where further research and development is not needed. In contrast, for technologies with potential for product development are protected by patents. The information disclosed in the patent provides knowledge about the Current State of Art, enables one to watch developments in industry assess the uniqueness of one’s own invention or find solution to specific technological problem.

In post genomic era Bioinformatics can be considered as central hub in drug discovery process. On support side of the hub is - information technology, information management, software applications databases and computational resources all provide infrastructure for bioinformatics. On scientific side of the hub are – application of bioinformatics methods in molecular biology, genomics, proteomics, and other emerging areas such as metabolomics, transcriptomics and in computer aided drug design research.

In our publication we have attempted to provide a wide coverage of patents on various aspects of drug discovery process which includes- genomic sequence, identification and validation of disease target, computer based screening of compounds, lead identification and lead optimization, in silico construction of drugs from component parts inside known molecular targets, and correlation between genotype and drug response, where bioinformatics plays a significant role.

We hope, the patent information included in the periodical will be useful for researchers in biomedical area. We shall be thankful to receive your comments/suggestion for improving this publication.

Thanks,

Dr. Sheela Tandon
Probabilistic method for determining nucleic acid coding features

US Patent No. 7,444,243
October 28, 2008

Abstract
The present invention is in the field of bioinformatics, particularly as it pertains to gene prediction. More specifically, the invention relates to the probabilistic analysis of nucleic acid sequences for the determination of coding features, including determination of state probabilities for each nucleotide in a nucleic acid sequence, determination of coding strand, determination of open reading frame extent, determination of insertion and deletion location, determination of exon location, and determination of protein sequence.

Bioinformatics system architecture with data and process integration

US Patent Application No. 20080033999
February 7, 2008

Abstract
A bioinformatics system and method is provided for integrated processing of biological data. According to one embodiment, the invention provides an interlocking series of target identification, target validation, lead identification, and lead optimization modules in a discovery platform oriented around specific components of the drug discovery process. The discovery platform of the invention utilizes genomic, proteomic, and other biological data stored in structured as well as unstructured databases. According to another embodiment, the invention provides overall platform/architecture with integration approach for searching and processing the data stored in the structured as well as unstructured databases. According to another embodiment, the invention provides a user interface, affording users the ability to access and process tasks for the drug discovery process.
Pathological tissue mapping

US Patent No. 7,483,554
January 27, 2009

Abstract
Embodiments of the present invention are directed to quantitative analysis of tissues enabling the measurement of objects and parameters of objects found in images of tissues including perimeter, area, and other metrics of such objects. Measurement results may be input into a relational database where they can be statistically analyzed and compared across studies. The measurement results may be used to create a pathological tissue map of a tissue image, to allow a pathologist to determine a pathological condition of the imaged tissue more quickly.

Information system for biological and life sciences research

US Patent Application 20090063259
March 5, 2009

Abstract
An online life science research environment and virtual community with a focus on design and analysis of biological experiments includes a life sciences laboratory system employing at least one networked computer system that defines a virtual research environment. Users access the system through a portal associated with the networked computer system(s). The virtual research environment has a data coupling mechanism by which the user designates a set of user-specified data for bioinformatics processing. A processor(s) associated with the networked computer system(s) performs bioinformatics services upon the user-specified data. In one embodiment, the data coupling mechanism enables transfer of the user-specified data to a memory space that is mediated or accessed by the processor performing the bioinformatics processing. Users may thus exploit bioinformatics processing resources that are not deployed on users' local computer environments, and to store and organize information relating to life sciences research in a secure, online workspace.
Abstract
Described herein is a system for inferring one or a population of biochemical interaction networks, including topology and chemical reaction rates and parameters, from dynamical or statical experimental data, with or without spatial localization information, and a database of possible interactions. Accordingly, the systems and methods described herein may be employed to infer the biochemical interaction networks that exist in a cell. To this end, the systems and methods described herein generate a plurality of possible candidate networks and then apply to these networks a forward simulation process to infer a network. Inferred networks may be analyzed via data fitting and other fitting criteria, to determine the likelihood that the network is correct. In this way, new and more complete models of cellular dynamics may be created.

Abstract
The present invention is in the field of bioinformatics, particularly as it pertains to gene prediction. More specifically, the invention relates to the probabilistic analysis of nucleic acid sequences for the determination of coding features, including determination of state probabilities for each nucleotide in a nucleic acid sequence, determination of coding strand, determination of open reading frame extent, determination of insertion and deletion location, determination of exon location, and determination of protein sequence.
Bioinformatics system architecture with data and process integration

US Patent Application No. 20080033999
February 7, 2008

Abstract
A bioinformatics system and method is provided for integrated processing of biological data. According to one embodiment, the invention provides an interlocking series of target identification, target validation, lead identification, and lead optimization modules in a discovery platform oriented around specific components of the drug discovery process. The discovery platform of the invention utilizes genomic, proteomic, and other biological data stored in structured as well as unstructured databases. According to another embodiment, the invention provides overall platform/architecture with integration approach for searching and processing the data stored in the structured as well as unstructured databases. According to another embodiment, the invention provides a user interface, affording users the ability to access and process tasks for the drug discovery process.

Streaming processing of biological sequence matching

US Patent No. 7,512,498
March 31, 2009

Abstract
A data system is provided for biological sequence matching. The system includes a system memory, a cache controller coupled to the system memory, a first cache coupled to the cache controller to receive non-temporal data from the system memory, and a second cache coupled to the cache controller to receive temporal data from the system memory. The first cache to also receive the temporal data from the second cache. The system further includes a processor coupled to the cache controller and the first cache.
**Snow mountain virus genome sequence, virus-like particles and methods of use**

US Patent No. **7,481,997**  
January 27, 2009

**Abstract**  
Snow Mountain Virus (SMV) belongs to the Norovirus genus of the Caliciviridae family. SMV is a genogroup II (GII) reference strain of human enteric caliciviruses associated with epidemic gastroenteritis. The positive sense RNA genome sequence of SMV was determined to be 7,537 nucleotides in length excluding the 3’ polyadenylated tract. The genome is organized into three open reading frames. Pairwise sequence alignments showed SMV ORF1 is highly conserved with other GII noroviruses, and most closely related to GII strains Melksham and Hawaii viruses. Comparative sequence analyses showed the SMV is a recombinant norovirus. VP1/NP2 proteins assembled into virus-like particles (VLPs) when expressed in insect cells by a recombinant baculovirus. Characterization of one clone that expressed VP1 but failed to assemble into VLPs, identified histidine residue 91 as important for particle assembly.

**Method for detecting DNA point mutations (single nucleotide polymorphism (SNP) analysis) and associated arrangement**

US Patent No. **7,488,578**  
February 10, 2009

**Abstract**  
The single nucleotide polymorphism analysis involves the utilization of a DNA hybridization process as well as the use of a DNA chip. A liquid DNA sample to be analyzed is guided over a DNA chip in a defined time course. After successful hybridization, the temperature is modified in a defined manner under low stringency conditions such that scavenger/target DNA hybrids are melted, whereby the melting of the scavenger/target DNA hybrids is detected and evaluated according to the temperature. In addition to the DNA chip, at least one device is provided that controls and regulates the temperature, and another device is provided that controls a lateral flow of liquid against the surface of the DNA chip. Factors for matching hybrids and mismatching/single point mismatching hybrids can be detected and evaluated using appropriate measuring device(s).
**URA5 gene and methods for stable genetic integration in yeast**

US Patent No. 7,514,253  
April 7, 2009

**Abstract**  
A novel gene encoding P. pastoris orotate-phosphoribosyl transferase (URA5) is disclosed. Methods for producing and selecting yeast strains capable of stable genetic integration of heterologous sequences into the host genome are also provided.

---

**Information System For Biological And Life Sciences Research**

US Patent Application No. 2009063259  
March 05, 2009

**Abstract:**  
An online life science research environment and virtual community with a focus on design and analysis of biological experiments includes a life sciences laboratory system employing at least one networked computer system that defines a virtual research environment. Users access the system through a portal associated with the networked computer system(s). The virtual research environment has a data coupling mechanism by which the user designates a set of user-specified data for bioinformatics processing. A processor(s) associated with the networked computer system(s) performs bioinformatics services upon the user-specified data. In one embodiment, the data coupling mechanism enables transfer of the user-specified data to a memory space that is mediated or accessed by the processor performing the bioinformatics processing. Users may thus exploit bioinformatics processing resources that are not deployed on users' local computer environments, and to store and organize information relating to life sciences research in a secure, online workspace.
System, method and computer program for non-binary sequence comparison

EP Patent No.2031533
March 4, 2009

Abstract
A system and method for performing non-binary comparison of biological sequences includes a new measure CÉ 0 , which is a non-binary counting measure that is used in a stand alone module called VaSSA-1. This measure obtains substantially more information about sequences and comparisons between them than is gathered by conventional bioinformatics techniques.

Biochip microsystem for bioinformatics recognition and analysis

US Patent Application No.2009048125
February 19, 2009

A system with applications in pattern recognition, or classification, of DNA assay samples. Because DNA reference and sample material in wells of an assay may be caused to fluoresce depending upon dye added to the material, the resulting light may be imaged onto an embodiment comprising an array of photodetectors and an adaptive neural network, with applications to DNA analysis. Other embodiments are described and claimed.
Method for analysis of pain images

Abstract
A method uses body images and computer hardware and software to collect and analyze clinical data in patients experiencing pain. Pain location information is obtained by the drawing of an outline of the pain on a paper copy or electronic display of the body image. Composite images are generated representing aggregate data for specified patient groups. The coordinates of common anatomic landmarks on differently designed body images are mapped to each other, permitting integrated analysis of pain data, e.g., pain shape, centroid, meta centroid, from multiple body image designs and display of all pain data on a single body image design. Differences and similarities between groups of patients are displayed visually and numerically, and are used to assign the probability of a given patient belonging to a particular diagnostic group or category of disease severity.

Data distribution method, data search method, and data search system

Abstract
Necessary information can be easily extracted from a plurality of databases in which biological substance information is stored. Data is downloaded from the plurality of databases. From the downloaded data is extracted information indicating links between two databases, a detailed description of each data, and sequence data for homology search, which together constitute an index. The thus extracted index is distributed to a user facility, where a user conducts a search using the distributed index.
Abstract
In one aspect, the invention provides a random or semirandom siRNA (encoding) library. Another aspect of the invention pertains to methods for construction of random or semirandom siRNA (encoding) libraries. Another aspect of the invention is vector systems for use in constructing siRNA libraries and/or that can express single siRNAs and siRNA libraries both constitutively and in an inducible fashion. In another aspect, the invention provides a method of using an siRNA library. The siRNA library is introduced into a population of cells. The population of cells then is subjected to a selection process to select a subpopulation of cells exhibiting a different behavioral, biochemical, chemical, functional, molecular, morphological, phenotypic, or physical property from the remainder of population. Following the selection process, the subpopulation of cells can be isolated, analyzed, and/or cloned as desired. Such analysis of the subpopulation can be identification and sequencing of the siRNA species responsible for the different properties of the subpopulation relative to the remainder of the population. Alternatively, the subpopulation can be further analyzed by genomic, proteomic, and/or cellomic assays. Where such genomic, proteomic, and/or cellomic assays are employed, the method can produce several useful bioinformatics products. Specific siRNAs identified through this process may have direct therapeutic value.

Viral Database Methods

Abstract
Disclosed are methods for designing oligonucleotides that can detect/identify any unknown or known virus of a particular taxon. Also provided are methods to establish, implement and validate bioinformatics tools and databases to support microarray design. The invention also provides specialized arrays for detection and speciation of select viral agents and viruses as well as a set of oligonucleotides that can detect/identify any unknown or known virus of a particular taxon.
Gene mining system and method

U S Patent No. 7,349,811
March 25, 2008

Abstract
The present invention provides a system, method and apparatus for targeting gene sequences having one or more phenotypic characteristics using a computer. One or more phenotypic characteristics are selected. A gene sequence is then selected that is known to have the selected phenotypic characteristics. In addition, one or more databases containing cataloged gene sequences are selected. The selected gene sequence is compared to the cataloged gene sequences, and any cataloged gene sequences that contain a portion of the selected gene sequence are extracted. The selected gene sequence is aligned to each portion of the extracted gene sequence and the extracted gene sequences are prioritized based on the alignment of the selected gene sequence. At least one of the prioritized gene sequences is selected based on one or more phenotypic criteria. Finally, one or more degenerate primers are designed to target the selected-prioritized gene sequences.

Methods for identifying DNA copy number changes

U S Patent No. 7,424,368
September 9, 2008

Abstract
Methods of estimating genomic DNA copy number are disclosed. Amplified genomic DNA is hybridized to an array of allele specific SNP probes to generate a hybridization pattern. A value, S, is calculated for individual SNPs in the experimental sample, where S is the log of the arithmetic average of the intensities of the perfect match probes for the SNP. S is calculated for the SNP in reference samples that are matched to the experimental sample in genotype. The mean and standard deviation for the S values of the reference samples are calculated and a log intensity difference is calculated by subtracting the mean values for the reference and experimental samples. The copy number of the SNP region is estimated using the difference between the mean for the SNP in the reference samples and the S value for the SNP in the experimental sample in a log-log linear model.
RECENT PATENTS ON BIOINFORMATICS IN DRUG RESEARCH

CHAIRMAN
Dr Tushar Kanti Chakraborty

ADVISORY BOARD
Dr. Madhan Mohan, DBT, New Delhi
Dr. G.P.S. Raghawan, IMT, Chandigarh
Dr. S, Krishanaswamy, MKU, Madurai
Dr. Indra Ghosh, JNU, New Delhi

EDITORIAL BOARD
Dr. Ravi Shankar
Dr. J V Pratap
Dr. M I Siddiqi

EDITORIAL TEAM
Sheela Tandon
Suman K Mallik

PUBLISHED BY: Scientist-in-charge, Knowledge Resource Centre & Coordinator, CDRI-BTIS Centre, Central Drug Research Institute, Chattar Manzil Palace, Lucknow-226001