

Bioinformatics up to Date

(Bioinformatics Center, Biotechnology Division)

North-East Institute of Science & Technology

Jorhat - 785 006, Assam



CONTENTS

COVER STORY

Cover Story	1
Bioinfy Quiz	1
Computational Chemistry	2
Proteomics	3
Computers for Biologists	3
Genomics	3
Bioserver	4
Software Mania	5
Molecule of the month	6
Bioinfy Animator	6
Contact Us	6
Answers to quiz	6
Upcoming Events	6

Adviser:

Dr.P.G.Rao

Editors:

Pompi Sharma

Dr. R.L.Bezbaruah

UC Irvine's Dr. Bang H. Hoang and colleagues have been presented the 2011 Kappa Delta Ann Doner Vaughan Award for their work advancing the understanding of how bone cancer spreads.

"Osteosarcoma is an aggressive cancerous bone tumor that appears primarily in children and is prone to spread from its primary location," said Hoang, a surgeon and researcher at UC Irvine's Department of Orthopaedic Surgery and Chao Family Comprehensive Cancer Center. "By understanding and explaining the underlying mechanisms for tumor progression, we hope eventually to develop therapies benefiting patients with sarcomas that are currently untreatable."

Though osteosarcoma can be treated in 60 percent to 70 percent of patients, the five-year survival rate for those who relapse is only 20 percent. With no significant change in bone cancer survival rates for the past two decades, Hoang sought to understand how inhibiting the Wnt pathway - a network of proteins linked to the progression of cancer - might suppress tumor growth and metastasis.

This work is reflected in his study, "Toward Novel Therapeutic Intervention for Osteosarcoma: Clinical Implications of the Wnt Pathway," co-authored by UC Irvine Drs. Yi Guo and Xiaolin Zi, as well as former UC Irvine Dr. Randall F. Holcombe.

<http://physorg.com>

Bioinfy Quiz

1) First bioinformatics database was created by:

- a) Dayhoff
- b) Durbin
- c) Pearson

2) Human genome contains:

- a) 2 billion base pairs
- b) 3 billion base pairs
- c) 4 billion base pairs

3) Analysing or comparing entire genome of species:

- a) Bioinformatics
- b) genomics
- c) Pharmacogenomics

4) Identification of drugs through genomic study:

- a) Chemi-informatics
- b) Pharmacogenetics
- c) Pharmacogenomics

5) Tool for identification of motifs:

- a) COPIA
- b) Patternhunter
- c) BLAST

Answers on Page 6

COMPUTATIONAL CHEMISTRY

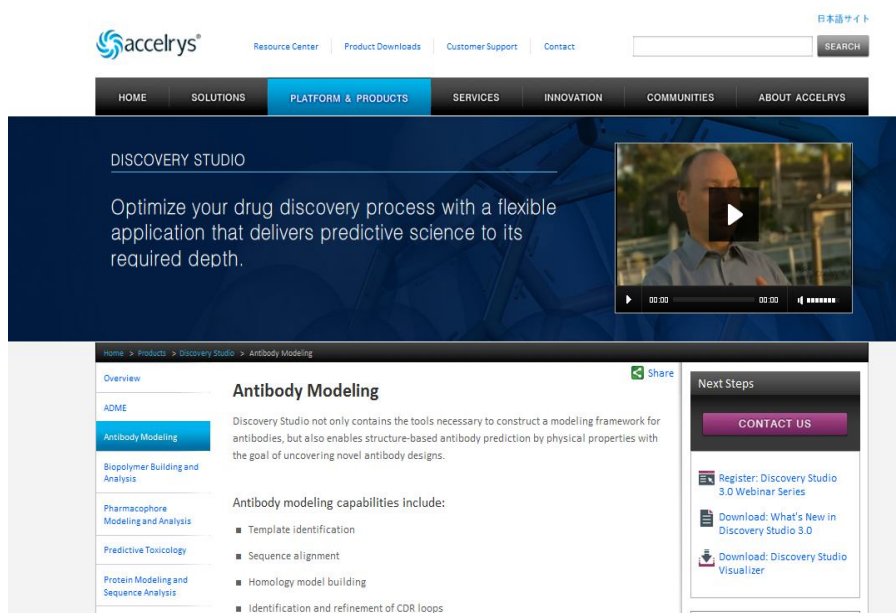
ACCELRY'S

Discovery studio

Discovery Studio 2.5.5 is software for computational chemistry and biology and is designed for applications in drug development. This product features algorithmic tools for protein modeling, access to the GOLD 4.1 docking program, tools to superimpose proteins, analyze binding site and build and edit nucleic acid. Protocols include a 3-D QSAR method and a protocol for calculating residue electrostatic energy for predicting protein stability and thermostability. The software also has two user interface Discovery Guides and 64-bit server compatibility.

Discovery Studio can:

- Investigate and test hypotheses in silico prior to costly experimental implementation, thus reducing the time and expense involved in bringing products to market.
- Drive scientific exploration from target identification to lead optimization with a wealth of trusted life science modeling and simulation tools.
- Leverage an open and scalable platform to automate processes, create and deploy custom workflows, and integrate data types, databases, and third-party or in-house tools.
- Enhance personal productivity and boost team collaboration by enabling researchers to share data and make better informed decisions.
- ADME, antibody modeling, biopolymer building and analysis, pharmacophore modeling and analysis, predictive toxicology, protein modeling and sequence analysis, QSAR and Library Design, simulations, structure based design, visualization, X-Ray.



The screenshot shows the Accelrys website interface. At the top, there is a navigation bar with links for Resource Center, Product Downloads, Customer Support, and Contact. A search bar is located on the right. Below the navigation bar, there is a main menu with options: HOME, SOLUTIONS, PLATFORM & PRODUCTS (highlighted), SERVICES, INNOVATION, COMMUNITIES, and ABOUT ACCELRY'S. The main content area features a large banner for DISCOVERY STUDIO with the text: "Optimize your drug discovery process with a flexible application that delivers predictive science to its required depth." A video player is embedded on the right side of the banner. Below the banner, there is a sidebar with a navigation menu for Discovery Studio, including Overview, ADME, Antibody Modeling (highlighted), Biopolymer Building and Analysis, Pharmacophore Modeling and Analysis, Predictive Toxicology, and Protein Modeling and Sequence Analysis. The main content area displays the "Antibody Modeling" section, which includes a description of the software's capabilities and a list of features: Template identification, Sequence alignment, Homology model building, and Identification and refinement of CDR loops. A "Next Steps" section on the right offers options to Register, Download, or Download the Visualizer.

<http://www.scientificcomputing.com/Products-DA-Discovery-Studio-032510.aspx>

PROTEOMICS

PROTEIN MUTANT DATABASE

The Protein Mutant Database (PMD) that we are constructing covers natural as well as artificial mutants, including random and site-directed ones, for all proteins except members of the globin and immunoglobulin families. The PMD is based on literature, not on proteins. That is, each entry in the database corresponds to one article which may describe one, several or a number of protein mutants. Each database entry is identified by a serial number and is defined as either natural or artificial, depending on the type of the mutation. For each entry the following items are recorded : "JOURNAL", "TITLE", "CROSS-REFERENCE", "PROTEIN", "N-TERMINAL", "CHANGE", "FUNCTION", "STRUCTURE", "STABILITY", etc. "CROSS-REFERENCE" indicates the code names of the protein given in other databases such as Protein Identification Resources (2). "N-TERMINAL" shows the N-terminal sequence of five amino acids which may help to show the unambiguous numbering of the sequence. "CHANGE" indicates the position and kind of mutations, such as amino acid substitution, insertion and deletion, denoted with a specific notation. Any functional or structural features ("FUNCTION", "STRUCTURE", "STABILITY", etc) observed in the mutant are described immediately after 'CHANGE'. Relative differences in activity and/or stability, in comparison with the wild-type protein, are indicated with symbols [- -],[-],[=],[+] or [+ +]. Complete loss of activity is denoted as [0]. It offers the options for both sequence search as well as keyword search.

<http://en.wikipedia.org> , <http://pmd.ddbj.nig.ac.jp>

COMPUTERS FOR BIOLOGISTS

Folding @ home

Folding @home ("Folding at Home") is a distributed computing (DC) project designed to perform computationally intensive simulations of protein folding and other molecular dynamics (MD), and to improve on the methods available to do so. It was launched on October 1, 2000. Folding @home does not rely on powerful supercomputers for its data processing; instead, the primary contributors to the Folding @home project are many hundreds of thousands of personal computer users who have installed a client program. The client runs in the background, utilizing otherwise unused CPU power.

The Folding @home client consists of three separate components:

- The client software acts as a download and file manager for work units and scientific cores, controls the cores, and is the software with which the user interacts. Separating the client from the core enables the scientific methods to be updated automatically (or new methods to be added) without a client update.
- The Work Unit is the actual data that the client is being asked to process.
- The Core performs the calculations on the work unit. Folding @home's cores are based on modified versions of seven molecular simulation programs for calculation: TINKER, GROMACS, AMBER, CPMD, SHARPEN, ProtoMol and Desmond.^{[11][12]} Where possible, optimizations are used to speed the process of calculation. <http://en.wikipedia.org>

GENOMICS

OMIM

OMIM- Online Mendelian Inheritance in Man. It is a database that catalogues all the known diseases with a genetic component, and—when possible—links them to the relevant genes in the human genome and provides references for further research and tools for genomic analysis of a catalogued gene. Every disease and gene is assigned a six digit number of which the first number classifies the method of inheritance. The information in this database was collected and processed under the leadership of Dr. Victor A. McKusick at Johns Hopkins University

Limits	Preview/Index	History	Clipboard	Details
---------------	----------------------	----------------	------------------	----------------

- To Search all fields, leave the following boxes unchecked.
- To narrow the search, check the boxes with specific fields' names, or use [search field tags](#) enclosed in square brackets, e.g. aaa[title].
- [Boolean operators](#) AND, OR, NOT must be in upper case.

Search in Field(s): clear	MIM Number Prefix: clear
<input type="checkbox"/> Title <input type="checkbox"/> MIM number <input type="checkbox"/> Allelic Variants	<input type="checkbox"/> * gene with known sequence
<input type="checkbox"/> Text <input type="checkbox"/> References <input type="checkbox"/> Clinical Synopsis	<input type="checkbox"/> + gene with known sequence and phenotype
<input type="checkbox"/> Gene Map Disorder <input type="checkbox"/> Contributors	<input type="checkbox"/> # phenotype description, molecular basis known
Chromosome(s): clear	<input type="checkbox"/> % mendelian phenotype or locus, molecular basis unknown
<input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8	<input type="checkbox"/> %? other, mainly phenotypes with suspected mendelian basis
<input type="checkbox"/> 9 <input type="checkbox"/> 10 <input type="checkbox"/> 11 <input type="checkbox"/> 12 <input type="checkbox"/> 13 <input type="checkbox"/> 14 <input type="checkbox"/> 15 <input type="checkbox"/> 16	Only Records with: clear
<input type="checkbox"/> 17 <input type="checkbox"/> 18 <input type="checkbox"/> 19 <input type="checkbox"/> 20 <input type="checkbox"/> 21 <input type="checkbox"/> 22 <input type="checkbox"/> X <input type="checkbox"/> Y	<input type="checkbox"/> Allelic Variants
<input type="checkbox"/> mitochondrial <input type="checkbox"/> unknown	<input type="checkbox"/> Clinical Synopsis
Creation Date <input type="text"/> From <input type="text"/> To <input type="text"/>	<input type="checkbox"/> Gene map locus
Last Modification <input type="text"/> From <input type="text"/> To <input type="text"/>	

Use the format YYYY/MM/DD; month and day are optional.

<http://en.wikipedia.org>

BIO SERVERS

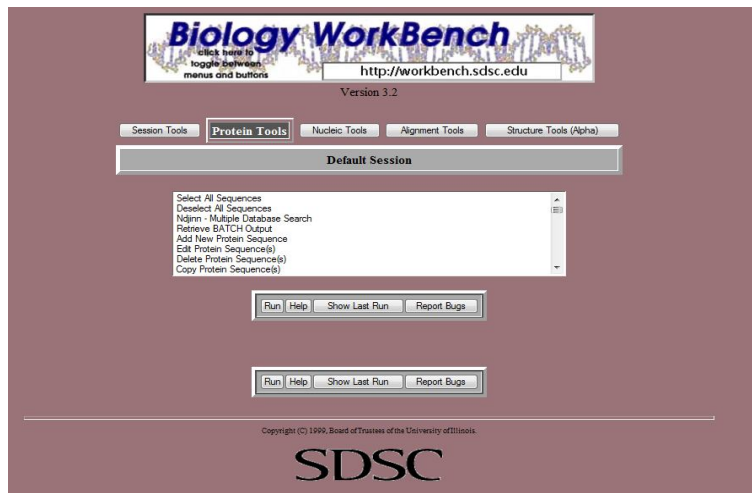
SDSC (San Diego Super Computer Centre)

BIOLOGY WORKBENCH

The Biology Work Bench is a web-based tool for biologists. The Work Bench allows biologists to search many popular protein and nucleic acid sequence databases. Database searching is integrated with access to a wide variety of analysis and modelling tools like: protein tools, session tools, nucleic tools, alignment tools, structure tools all within a point and click interface that eliminate file format compatibility problems. It was originally developed for Netscape Communicator or Navigator, up through version 4.7x.

Features of the biology workbench:

- Full support for modelling and visualization of biological structures, including an integrated tool (Sirius)
- The ability to save and view input, parameter, and output files for all jobs that are run
- New phylogenetic tree inference tools (from Phylip)
- Multiple clickable folders for organizing projects
- MFold for RNA structure prediction



<http://workbench.sdsc.edu>

SOFTWARE MANIA

SOSUI

SOSUI is a free online tool that predicts a part of the secondary structure of proteins from a given amino acid sequence (AAS). The main objective is to determine whether the protein in question is a soluble or a transmembrane protein. SOSUI's algorithm was developed in 1996 at Tokyo University. The name means as much as "hydrophobic", an allusion to its molecular "clients".

How SOSUI works

First of all, SOSUI looks for α helices that are relatively easy to predict, taking into account the known helical potentials of the given amino acid sequence(AAS). SOSUI uses 4 characteristics of the AAS in its prediction:

1. "hydropathy index" (Kyte und Doolittle 1982)
2. weighted presence of amphiphilic amino acids (AA) and their localization: "amphiphilicity index"
3. the AA's charge
4. the length of the AAS

Results

The result page first shows general information (length, average hydrophobicity). If the protein in question is a transmembrane protein, the number of transmembrane domains and their localization is noted. A "hydropathy-profile" with colored accentuation of hydrophobic parts; the helical wheel diagrams of potential transmembrane domains are shown as well.

<http://en.wikipedia.org>

http://bp.nuap.nagoya-u.ac.jp/sosui/sosui_submit.html

Molecule of the month

INTEGRASE

Integrase is the enzyme that splices the viral DNA into a cellular chromosome. Four identical copies of integrase grab the two ends of the viral DNA, creating a stable complex called an intasome. The intasome then binds to the cellular DNA and performs a strand transfer reaction, joining the viral DNA to the cellular DNA.

Molecular Data

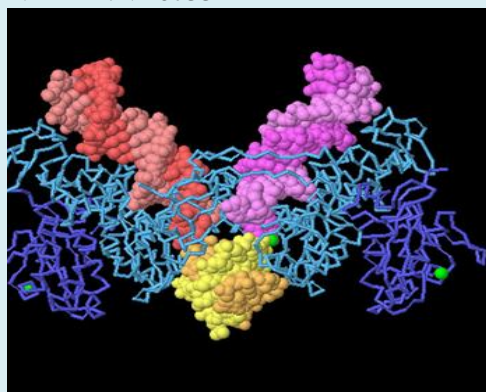
PDB ID: 3os1

Structural Weight: 109365.71

Length: 395

Method: X-Ray Diffraction

NDB ID: NA0755



www.rcsb.org

UPCOMING EVENTS

3 days National workshop

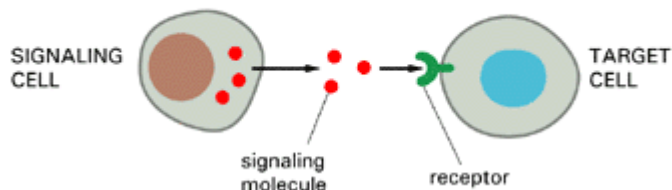
On “**Structural bioinformatics tools and application**” from 23rd to 25th April 11 @ Bio intelligence

Office, Indore

e-mail: info@biointelligence.in

CELL SIGNALING

SIGNALING BY SECRETED MOLECULES



SIGNALING BY PLASMA-MEMBRANE-BOUND MOLECULES

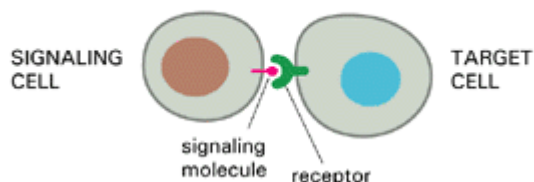


Fig: Diagram showing the two forms of cell signaling: chemical signaling and receptor-mediated signaling.

www.google.co.in

Answers to quiz: 1)a; 2)b; 3)b; 4)c;
5)a

CONTACT US:

Pompi Sharma, Project Assistant (Level – II)

E-mail: pompi.sharma86@gmail.com