



Bioinformatics up to Date

(Bioinformatics Center, Biotechnology Division)
North-East Institute of Science & Technology
Jorhat - 785 006, Assam
(<http://www.rrljorhat.res.in/biotechnology.html>)

CONTENTS

Cover Story	1
Bioinfo Career	1
Computational Chemistry	2
Proteomics	2
Bioserver	3
Computers for Biologists	3
Genomics	4
Upcoming Events	4
Structure Bioinfo.	5
Molecule of the month	5
Contact Us	5

Adviser:
Dr.P.G.Rao

Editors:
Dr. R.L.Bezbaruah
Mr Dhrubjyoti Gogoi

**Bioinformatics
Centre of CSIR
-NEIST, Jorhat
wishing you a
very happy
New Year, 2012**

COVER STORY

Environmental Niche Modelling A discipline of biodiversity informatics

Environmental niche modelling, alternatively known as species distribution modelling, (ecological) niche modelling, and climate envelope modelling refers to the process of using computer algorithms to predict the distribution of species in geographic space on the basis of a mathematical representation of their known distribution in environmental space (= realized ecological niche). The environment is in most cases represent by climate data (such as temperature, and precipitation), but other variables such as soil type, water depth, and land cover can also be used. These models allow for interpolating between a limited number of species occurrence and they are used in several research areas in conservation biology, ecology and evolution.

The extent to which such modelled data reflect real -world species distributions will depend on a number of factors, including the nature, complexity, and accuracy of the models used and the quality of the available environmental data layers; the availability of sufficient and reliable species distribution data as model input; and the influence of various factors such as barriers to dispersal, geological history, or biotic interactions, that increase the difference between the realized niche and the fundamental niche. Environmental niche modelling may be considered a part of the discipline of biodiversity informatics.

Niche modelling software :

- 1.Maxent software for species habitat modeling
- 2.ModEco: Integrated Software for Species Distribution Analysis and Modeling
- 3.DIVA-GIS

(Ref: <http://www.esajournals.org/doi/abs/10.1890/08-0134.1>)



As of Tuesday Dec 20,
2011 at 4 PM PST
there are **78020**
Structures

BIOINFO. CAREER

Scientist job: @NEIST
Jorhat.

No. of Post: One, Area:
Bioinformatics,
Deptt./Division: Biotechno-
logy, Reservation: UR
Essential Qualifications:
PhD in Bioinformatics
(Thesis submitted/
awarded)
Details Add: refer: NEIST
website.

Walk-in-interview for
Bioinformatics JRF at
Directorate of Wheat Re-
search, Karnal
DBT Funded Project on
"Phenotyping of mapping
populations at hot spots
and tagging of major
QTLs associated
with spot blotch resistance
in wheat" ,Agri bio-
informatics promotion
programme,([http://dwr.in/
images/pdf/11/
srfjrfpostsad-
vdec2011.pdf](http://dwr.in/images/pdf/11/srfjrfpostsad-vdec2011.pdf))

Biotech Desk Pvt. Ltd. offers
career opportunities in bio-
technology and related fields
(www.biotechdesk.com/)

Hash BioTech Labs Private
Limited conducts functional
training programs in the field
of Micro Biology and Bio-
technology.
(www.hashbiotech.com)

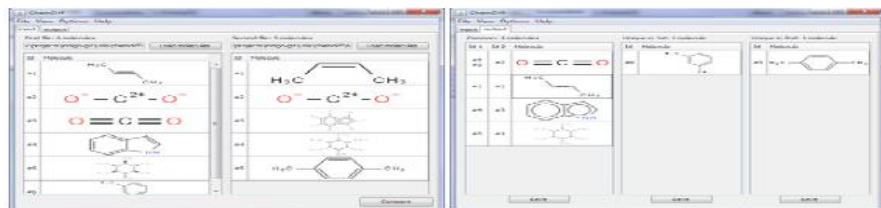
Computational Chemistry

ChemDiff

ChemDiff is an Indigo-based utility for finding duplications and visual comparison of two files containing multiple structures. SDF, SMILES, CML, MOLFILE input formats are supported. Files can contain large amount of molecules and ChemDiff was tested on files with up to 1 million ones.

There are at least 3 possible use cases for ChemDiff:

1. Compare 2 sets of molecules.
2. Find duplicates in a single set of molecules.
3. Test whether a specified molecule (or some set of molecules) belongs to another set of molecules.



ChemDiff window contains two tabs: first tab is for input molecules, and second tab contains comparison result.

Proteomics SimiCon

SimiCon will identify the equivalent protein-ligand atomic contacts (EC) between Reference (RC) and Target (TC) complexes. The results will be shown as text, tables and 3D interactive graphics.

Reference and target complexes can be uploaded as PDB coordinate files, or retrieved by using PDB IDs.

- PDB ID, according to Protein Data Bank nomenclature.
- Chain, the chain selection must correspond to Protein Data Bank nomenclature. If the form is left blank, the first chain will be processed. Note that only single chains are allowed.
- Ligand, the name must correspond to the PDB nomenclature. If the form is left blank, the first HETATM molecule will be processed.

In "Upload mode", only the first protein chain and the first HETATM molecule will be analyzed. If you want to include a "multi-chain" receptor (or ligand) in a single run, please join the chains manually.

- Distance cutoff, the recommended standard cutoff distance for heavy-atom contacts is 4 Å, but the user can choose a range from 2 to 12 Å.

The server allows to compare two complexes having non-identical protein sequences, however, only aligned amino acids will be analyzed and displayed. If your sequences differ, you might want to try performing "self" calculations (e.g., the same complex being reference and target) to retrieve the actual contacts for both complexes.

The server is now capable of handling different ligands. In such a case, the results for the target will only be shown for the maximal common chemical substructure. This feature is provided for users trying to compare "slightly" different ligands. If your ligands do not have any common substructure, then a simply "receptor-based" contact measure (i.e., counting which common receptors atoms have contact/s with any ligand heavy-atom/s) may be more appropriate.

Bioserver

Traditional Chinese Medicine Database@Taiwan

Traditional Chinese Medicine Database of Taiwan is currently the most comprehensive and largest non-commercial TCM database available for download. All the molecular file formats provided in the database can be readily used for docking and molecular dynamics simulation. Right now, the database contains constituents from 352 different herbs, animal products and minerals. Some of the key features of this database are:

The database contains 37,170 (32,364 non-duplicate) TCM compounds from 352 TCM ingredients. Users may search the database based on different criteria, including chemical name, TCM name, molecular properties and molecular structure.

The molecules in the database are available for download in cdx (2D structure) and mol2 (3D structure) file format. All molecules have been geometrically minimized and could be used in docking programs.

Unlike other traditional medicine website or database, the Traditional Chinese Medicine Database of Taiwan, provides comprehensive ingredient structural information of each TCM. In addition, original research articles of each ingredient are available in reference section.

Users may download TCM subsets or desired TCM ingredients. Currently, the entire database may be downloaded as well.

The Traditional Chinese Medicine Database of Taiwan is created in the vision of an open TCM network that is free for researchers to share their data. Thus, users are free to upload their own molecules and the associated reference works. In this way, the Traditional Chinese Medicine Database@Taiwan can continue to grow and keep the database information most up-to-date.

Currently, only TCM ingredients are available. In the future, we wish to extend the database to include traditional folk medicine as well.

As technology advances, new TCM ingredients could be isolated in a daily basis. In the hope to create the world largest and most complete TCM library and network, we would like to call out for research groups to share their TCM findings with the society.

Computers for Biologists

DIVA-GIS

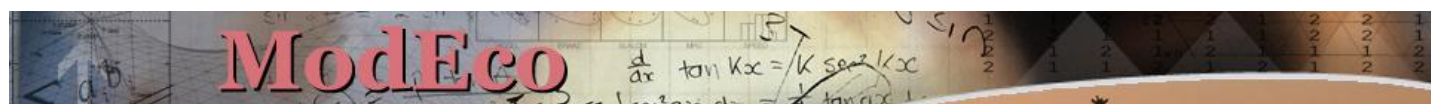
DIVA-GIS is a free computer program for mapping and geographic data analysis (a geographic information system (GIS)). With DIVA-GIS you can make maps of the world, or of a very small area, using, for example, state boundaries, rivers, a satellite image, and the locations of sites where an animal species was observed. We also provide free spatial data for the whole world that you can use in DIVA-GIS or other programs.

You can use the discussion forum to ask questions, report problems, or make suggestions. Or contact us, and read the blog entries for the latest news. But first download the program and read the documentation.

DIVA-GIS is particularly useful for mapping and analyzing biodiversity data, such as the distribution of species, or other 'point-distributions'. It reads and write standard data formats such as ESRI shapefiles, so interoperability is not a problem. DIVA-GIS runs on Windows and (with minor effort) on Mac OSX (see instructions).

ModEco: Integrated Software for Species Distribution Analysis and Modeling

ModEco is a software package for ecological niche modeling. It integrates a range of niche models within a geographical information system. ModEco provides a user friendly platform that enables users to explore, analyze, and model species distribution data with relative ease. It contains a suite of functions such as data input/output, data visualization, feature analysis, model training and prediction, and accuracy assessment, which are commonly



Genomics

GENECLASS2**A Software for Genetic Assignment and First-Generation Migrant Detection**

GENECLASS2 is a software that computes various genetic assignment criteria to assign or exclude reference populations as the origin of diploid or haploid individuals, as well as of groups of individuals, on the basis of multilocus genotype data. In addition to traditional assignment aims, the program allows the specific task of first-generation migrant detection. It includes several Monte Carlo resampling algorithms that compute for each individual its probability of belonging to each reference population or to be a resident (i.e., not a first-generation migrant) in the population where it was sampled. A user-friendly interface facilitates the treatment of large datasets. Three types of criteria used for likelihood estimation have been implemented in GENECLASS2: genetic distance-based criteria, a criterion directly based on allele frequencies, and Bayesian criteria

Running Environment

GENECLASS2 was developed in the Pascal object programming language and compiled with Borland Delphi6 and Kylix2. Therefore the software can be run on a Microsoft Windows or a Linux platform. An easy-to-use graphics interface has been designed to guide the user in the assignment process: choice of task (assign/exclude population as origin of individuals or detection of migrants), choice of statistical criterion for likelihood estimation, computation of probabilities, etc. The package includes a user-friendly help file with graphical interfaces that explain how to run the program to perform the two above tasks.

Program Availability

GENECLASS2 is freely available in English or French at <http://www.montpellier.inra.fr/CBGP/software>. A self-extracting setup executable leads the user through installation of the software on Windows-based machines. An RPM file containing the program file and the libraries allows the package to be installed on Mandrake and Red Hat Linux platforms. A .tar.Z archive allows manual installation of the binaries on other kinds of Linux platforms. A registration form allows users to be kept informed of new releases.

Upcoming Events

18TH INTERNATIONAL CONFERENCE 2012 (POST ISCBC)

Perspective and Challenges in
Chemical and Biological Sciences

Innovation Cross Roads

28th - 30th JANUARY, 2012

JOINTLY ORGANIZED BY

INSTITUTE OF ADVANCED STUDY IN SCIENCE AND TECHNOLOGY (IASST), GUWAHATI,
ASSAM, INDIA

AND

INDIAN SOCIETY OF CHEMISTS AND BIOLOGISTS (ISCB), LUCKNOW, UP, INDIA

www.iasst.in www.iscbindia.com

INSTITUTE OF ADVANCED STUDY IN SCIENCE AND
TECHNOLOGY (IASST)

(An Autonomous Institute under Department of Science and
Technology, Govt. of India)

Vigyan Path, Paschim Boragaon, Garchuk, NH-37, Guwahati - 781
035

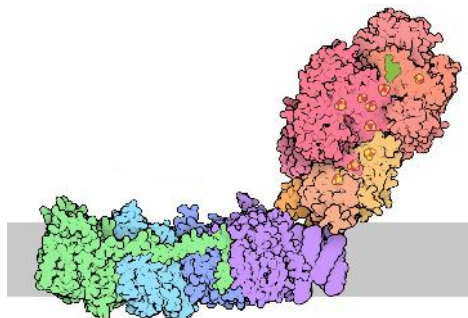


Venue:

[View Map](#)

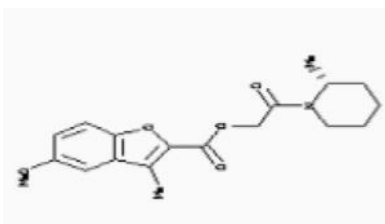
Structure Bioinfo.

Complex I hydrophilic domain PDB ID: 2FUG



Complex I, also known as NADH:quinone oxidoreductase, performs the first step in respiratory electron transport. This structure includes the hydrophilic domain, which performs the electron transport reaction.
(Ref:<http://www.rcsb.org/pdb/101/motm.do?momID=144>)

Molecule of the month



SMILES: Cc1c2cc(ccc2oc1C(=O)OCC(=O)N3CCCC[C@H]3C)OC

Physical Representations

pH representation	xlogP	Apolar desolvation (kcal/mol)	Polar desolvation (kcal/mol)	H-bond donors
Reference (pH 7)	3.33	2.09	-12.42	0
H-bond acceptors	Net charge	tPSA (Å ²)	Molecular weight (g/mol)	Rotatable bonds
6	0	68	345.395	5

REF: <http://zinc.docking.org/substance/3365133>

Kindly send us your feedback to

Dhrubajyoti Gogoi,

Editor, Bioinformatics up to date and Project Assistant (II),

Biotechnology Division, CSIR-North East Institute of Science and Technology, Jorhat, Assam

E-mail: dhruba.bio.du@gmail.com

Mobile: 7896689444(M)