

# Book of Abstracts\*

## Inbix 10

12th - 13th February 2010

Indian virtual Bioinformatics conference 2010

Supported by the Bioinformatics Organization & Asia Pacific Bioinformatics Network



---

\* Although the abstracts are proof read, readers' discretion is advised. Inbix'10 organizers do NOT claim any responsibility for technical editing

## Table of Contents

Serial Number	Paper Code	Title	Page
1	Inbix10/3	Riboswitches: Novel Drug Target for Substitution of Antibiotics in Tuberculosis	3
2	Inbix10/5	Computational prediction of proteome separated by 2DE in <i>Phanerochaete chrysosporium</i>	4
3	Inbix10/6	Base composition and tetra nucleotide motifs and their repeats in plastid genomes	5
4	Inbix10/7	Identification of molecular marker for Prolamin seed storage protein in millets	6
5	Inbix10/8	Prediction of Oral Bioavailability of drugs by Machine Learning approach	7
6	Inbix10/9	Homology modelling of Salmonella Typhi Ty2 enolase and its docking studies	8
7	Inbix10/10	Comparative study of seed storage protein sequences in cereals	9
8	Inbix10/11	In-silico toxicity prediction of ligand molecules by Support Vector Machine	10
9	Inbix10/12	Transcription factor binding site database	11
10	Inbix10/13	In silico prediction of T-cell epitopes from outer membrane protein of outer membrane vesicles of <i>Neisseria meningitidis</i> serogroup B (MC58) for vaccine design	12
11	Inbix10/14	Comparative analysis of computational epitope predictions: proposed library of putative vaccine candidates for HIV	13

### Inbix10/3

#### **Riboswitches: Novel Drug Target for Substitution of Antibiotics in Tuberculosis**

Neel Mehta\*, Prashanna Balaji

Third Year B.Tech Bioinformatics  
Maulana Azad National Institute of Technology, Bhopal – 462007, India

\*Corresponding author: [neel1289@gmail.com](mailto:neel1289@gmail.com)

#### **Abstract:**

Tuberculosis is a killer bacterial disease that destroys a whopping three million annually. Disease is caused by *Mycobacterium tuberculosis* bacteria that primarily affect the lungs in humans. The disease is treated using various antibiotic drugs, which is to be consumed for a long time resulting into its ineffectiveness against bacteria that show multiple drug resistance. Therefore, there occurs a strong need to look for other potent sources; Riboswitches seems to be an optimistic solution: 1. To Predict Regulatory RNA's such as Riboswitches within *Mycobacterium tuberculosis* bacteria for curing Tuberculosis. 2. Identification of unique folding patterns in Riboswitches. Riboswitches are highly structured cis-acting elements located in the 5'-untranslated region (UTR) of mRNA that directly binds small molecule metabolites to regulate gene expression by feedback inhibition. Riboswitches are divided in two parts: an aptamer and an expression platform. The aptamer directly binds small molecule and expression platform undergoes structural changes in response to changes in aptamer. A tool *viz* RibEx: Riboswitch Explorer was implemented for identifying Riboswitches, a part of the genome of *Mycobacterium tuberculosis* was taken from the NCBI genome database and BLISS Database. Mainly two Riboswitches, Coenzyme B12 or Adenosylcobalamin (AdoCbl) and Thiamine Pyrophosphate (TPP) were found after sequence analyses.

**Keywords:** Riboswitches, database

**Inbix10/5****Computational prediction of proteome separated by 2DE in *Phanerochaete chrysosporium***

Sivaramaiah Nallapeta\*, Ranganath Gudimella and Krishna Mohan  
Birla Institute of Scientific Research, Jaipur 302001, Rajasthan, India

\* Corresponding author: [myresearchbin@gmail.com](mailto:myresearchbin@gmail.com)

**Abstract**

Two-dimensional gel electrophoresis (2D-PAGE) is one of the most comprehensive techniques for separating thousands of proteins simultaneously with high reproducibility. The completion of numerous genome sequencing projects offers the rapid identification of these protein spots via N-terminal amino acid sequence determination or mass spectrometry in combination with peptide mass fingerprinting. The complex pattern of protein spots usually observed is mainly the result of the individual separation behavior of each protein dependent on its unique molecular weight (Mw) and isoelectric point (pI). However, the degree of visibility of individual proteins of the pattern relies on their cellular concentration, the amount of closely separating proteins and the separation quality of the chosen pH/Mw range of the electrophoretic system. Calculation of the theoretical Mw and the pI from the amino acid composition of a protein in combination with the physical principles of the employed electrophoretic separation procedures, allows determination of the approximate position of protein spots during 2D PAGE. After determination of these two parameters it is possible to construct and visualize a virtual 2D gel. This may serve both as a reference map for protein spot identification and for the choice of the optimal pH gradient in order to characterize proteins of interest. Hence in our current study reference maps for proteome separated by 2DE were constructed and finally compared with *in-silico* protein database of *P. chrysosporium* to characterize our proteins of interest.

**Keywords:** 2D-PAGE, Proteomics, reference maps.

**Inbix10/6****Base composition and tetra nucleotide motifs and their repeats in plastid genomes**Riju A<sup>1</sup>, Malhotra V<sup>2</sup> and Arunachalam V<sup>2</sup> \*

1Aikkal, Kanul PO, Kannur, Kerala 670 564, India, 2 Senior Software Engineer, Infogain Pvt. Ltd. Noida, India. 3 Present address Bioinformatics Centre, Marikunnu, Calicut, Kerala, India.

\*Biotechnology Laboratory, Division of Crop Improvement, Central Plantation Crops Research Institute, Kudlu Post, Kasaragod 671124, Kerala India, Phone: +91-4994-232894 Fax: +91-4994-232322

\* Corresponding author: [vadivelarunachalam@yahoo.com](mailto:vadivelarunachalam@yahoo.com)

**Abstract:**

Plastids are important organelle unique to land plants, algae and protozoa. Our objective is to analyze the composition of four nucleotides and density of five tetra nucleotide motifs (TATA, CAAT, GATA, GACA and CACA) and their repeats in each of the 107 complete plastid genomes. Perl scripts were written for the purpose and used. Plastomes are rich in Adenine and Thyime (A+T) ranging from 57.9 % to 80.5 %. Plastid genomes are rich in TATA and its repeats but poor in GACA / CACA or its repeats. TATA motifs occur at high density 7.19 (2.98 to 24.09) per Kb in plastid genomes. About 32 out of 107 genomes had quadruplet repeats and 13 of them had pentameric repeats of TATA. CACA motifs and GACA motifs are the low dense motifs in plastomes at 2.04 (1.12 to 3.65) per Kb and 2.17 (0.857 to 2.78) per Kb respectively. Lowest density of motifs CACA, GACA and highest density of TATA are seen in apicoplasts. Zygnema genome has highest density of CAAT and CACA motifs. Tetramer densities in 21 sample genomes of varying size, GC content and taxonomy were analysed using a Z score statistics and correlation coefficients. Significant Z score for all five tetramer densities was seen only in Spinacia genome alone. GACA density was significantly overrepresented in ten genomes. Taxonomically proximal plastomes are highly correlated and vice versa. The results are freely available at <http://riju.byethost31.com/plastids/>. Results are discussed in relation to evolution, DNA bending, RNA editing and nucleosome signalling.

**Keywords:** Chloroplast, TATA, nucleotide, sequence, DNA winding, CACA, GATA, CAAT, GACA.

**Inbix10/7**

**Identification of molecular marker for Prolamin seed storage protein in millets**

A.N. Nayak<sup>1</sup>, Rahul Kumar<sup>1</sup>, Sharad Panday<sup>1</sup>, Sudepto Bhattacharya<sup>2</sup> and Meera Narasimha<sup>1\*</sup>

<sup>1</sup>Department of Biotechnology, Faculty of Science & Technology, ICFAI University, Dehradun  
248197, Uttarakhand, India

<sup>2</sup>Department of Mathematics, Faculty of Science & Technology, ICFAI University, Dehradun  
248197, Uttarakhand, India

\*Corresponding author: [meera.narasimha@gmail.com](mailto:meera.narasimha@gmail.com)

**Abstract**

Prolamins are the major seed storage proteins in most cereal seeds and accumulate to high levels in seeds during the late stages of seed development. Apart from storage, prolamins have no known function. Although all are soluble in aqueous alcohols, they vary considerably in their other properties. Maize, sorghum and pearl millet contain similar groups of prolamins. Zein, kafirin and pennisetin are the alcohol soluble prolamins in maize, sorghum and pearl millet respectively. Sequences of zein, kafirin and pennisetin were retrieved from the GenBank. The sequences retrieved were aligned to each other using BLAST for sequence similarity wherein we found that the sequences have considerable homology and further analyses through multiple sequence alignment was carried out to find out the conserved region in three major prolamins of different group. Primique online prifi-primer design tool was used to find out the primer for the queried sequences. We found that 72 to 78% identities were aligned and 5-9% gaps were found indicating that the considerable homology is present amongst the gene sequences of zein, pennisetin and kafirin. Furthermore 32 primer pairs were designed for the three query seed storage prolamine gene sequences. However, their validation is required to be confirmed before use them as molecular markers in future.

**Keywords:** Prolamins, molecular markers

**Inbix10/8****Prediction of Oral Bioavailability of drugs by Machine Learning approach**

Anju Sharma and Rajnish Kumar\*

Department of Bioinformatics, Indian Institute of Information Technology Allahabad

\*Corresponding author: [12.rajnish@gmail.com](mailto:12.rajnish@gmail.com)**Abstract:**

Oral Bioavailability is defined as the rate and extent to which an active drug substance is absorbed and becomes available to the general circulation. An *in silico* model for predicting oral bioavailability is very important both in the early stage of drug discovery to select the promising compounds for further optimizations and in later stage to identify candidates for clinical trials. Various experimental and in-silico methods were being used in past for the prediction of oral bioavailability of drugs. In present study we have proposed a Support vector Machine based kernel learning approach carried out on a set of 511 chemically diverse compounds with known oral bioavailability values. For each drug 12 descriptors were calculated. Optimal values of C and  $\gamma$  were determined by running grid parameter search with 384 training set data and the prediction efficiency of proposed classifier was tested on remaining 127 test set data. The overall prediction efficiency for test set came out to be 96.85%. Youden's index and Matthew correlation index were found to be 0.929 and 0.909 respectively. Area under ROC curve was found to be 0.943 with standard error 0.0253. This prediction accuracy suggests that SVM-based prediction of oral bioavailability can be considered as helpful tool in drug discovery and development.

**Keywords:** Oral Bioavailability, Support Vector Machine**Acknowledgments:** We thank Dr. Pritish Varadwaj for his guidance on this work

**Inbix10/9****Homology modelling of *Salmonella Typhi Ty2* enolase and its docking studies**

Digvijay Singh\* and Sharat Chandra,

School of Biotechnology, Devi Ahilya University, Indore, India

\*Corresponding author: [chauhantigers@gmail.com](mailto:chauhantigers@gmail.com)**Abstract:**

Enolase an enzyme, of *Salmonella enterica serovar Typhi Ty2* is a human specific pathogen an aetiological agent for typhoid fever. It has ability to spread deeper into tissues of humans including liver, spleen and bone marrow. Enolase is a glycolytic enzyme, belongs to a novel class of surface proteins and is a multifunctional protein having ability to serve as plasminogen receptor. It is an established fact that the enzyme inhibition requires structural determinants from its 3D structure, instead of sequence of amino acids. In our present study, molecular modelling of enolase salmonella is achieved *in silico* by comparative modelling. MODELLER program was adopted for homology modelling which provides an accurate and efficient module to build model with identical structures. On the basis of Ramachandran plot, in the present study conformational parameters  $\Phi$  and  $\Psi$  angles were calculated from model with 94.7% residues in most favoured region. Further, PROCHECK results confirmed acceptance of model through main and side-chain values. We believe that structure can be an important tool to understand the interaction between a pathogen and the host, which may provide some novel potential therapeutic targets by exploring various inhibitors for the studied enzyme. However, many recent reports based on the molecular docking analyses between enolase salmonella and human plasminogen suggested the direct physical interaction between both of them. From the present study, we suggest that theoretical comparison for the similar structures of different enolase salmonella, resistant to multiple drugs, may be helpful for the better understanding of the invasion mechanisms.

**Keywords:** Homology modeling, Glycolytic enzyme, MODELLER, PROCHECK**Acknowledgement:** We thank Dr. Tiratha Raj Singh for scientific support.

**Inbix10/10****Comparative study of seed storage protein sequences in cereals**

Rahul Kumar<sup>1\*</sup>, Poonam R. Pal<sup>1</sup>, Sudepto Bhattacharya<sup>2</sup>, Sharad Panday<sup>1</sup> and Meera Narasimha<sup>1</sup>

<sup>1</sup>Department of Biotechnology, Faculty of Science & Technology, ICFAI University,  
Dehradun 248197, Uttarakhand, India

<sup>2</sup>Department of Mathematics, Faculty of Science & Technology, ICFAI University,  
Dehradun 248197, Uttarakhand, India

\*Corresponding author: [rahuldehran007@rediffmail.com](mailto:rahuldehran007@rediffmail.com)

**Abstract :**

Cereals have two important classes of storage proteins *viz* globulins and prolamins. Seed storage protein from wheat, rye, barley, oat and maize were used to find out the evolutionary distance. Prolamins show more variable in structure than globulins. Using the available databases, the similarity of different cereal proteins were analyzed using BLAST, multiple sequence alignment and ClustalW tools. The BLAST analyses of proteins showed some similarity in amino residue sequences in storage proteins. Furthermore, Secalin of rye was known to be aligned with Avenin of oat indicating 72% of positives by local alignment even as we found that the Glutamine repeat (QQ) sequences were found in many regions. Protein sequence of zein of maize and gliadin of wheat aligned with each other showing 44% identity and 66% true positives. Gliadin of wheat and hordein of barley showed 85% identities and 85 true positives while Secalin and gliadin showed 78% and 82% of identity, positives respectively. We found that Leucine and proline sequence repeats (LP) were found in these regions suggesting that these exhibit tandem repeats and th repetitive peptides were slightly different from those of prolamins of other families. In turn, these repeats suggest that they might play a role in formation of helix structure. Furthermore phylogram using ClustalW resulted in alignment of seven different proteins which are: alpha-avenin, alpha-gliadin, omega-secalin, D-hordein, gamma-secalin, alpha-zein, gamma hordien and evolutionary distances were compared amongst them. We conclude that Alpha-zein with a large distance (0.47292) could throw light on a separate evolutionary origin than that of prolamine present in oat, wheat, rye and barley.

**Keywords:** Prolamins, Evolution

**Inbix10/11**

**In-silico toxicity prediction of ligand molecules by Support Vector Machine**

Rajnish Kumar and Anju Sharma\*

Department of Bioinformatics, Indian Institute of Information Technology Allahabad

\*Corresponding author: [anjusharma.online@gmail.com](mailto:anjusharma.online@gmail.com)

**Abstract:**

The in-vivo or in-vitro prediction of toxicity is time consuming and an expensive process and plays a significant role in the late-stage drug failure. Efforts are being constantly made to make it possible to accurately predict the in-vivo toxicity of compounds directly from their structure without carrying out the in-vitro studies of compounds. Various in-silico toxicity prediction models have been developed. Although a number of prediction methods have been reported, still there is room for improvement owing to the fact that the ability of the prediction model depends on the type of descriptors they are using to represent a molecule. In this study, we presented RBF kernel based SVM classification model for the toxicity prediction of known as well as unknown compounds, using 17 descriptors. The selection of optimal hyperplane parameters were performed with 1696 training ligand data and the prediction efficiency of proposed classifiers were tested on remaining 566 test data. The overall prediction efficiency for test set came out to be 70.50%. Youden's index and Matthew correlation index were found to be 0.42 and 0.44 respectively. Area under ROC curve was found to be 0.755 with standard error 0.306. The overall performance of the model is satisfactory, equivalent to other reported methods and needs further enhancement.

**Keywords:** Toxicity prediction, Support Vector Machine

**Acknowledgments:** We thank Dr. Pritish Varadwaj for his guidance on this work.

**Inbix 10/12****Transcription factor binding site database**Mohit Sharma\*<sup>1</sup>, Manoj Dhar<sup>1</sup> and Vladimir B. Bajic<sup>2</sup>

1 Bioinformatics Centre, School of Biotechnology, Jammu University, Jammu 180006, India

2 A-star Singapore, Singapore

\* Corresponding author: [ksmohit@gmail.com](mailto:ksmohit@gmail.com)**Abstract:**

Transcription factors are key regulatory proteins for controlling the transcription activation in every organism. These proteins bind to a specific DNA sequence which will result in either an increasing rate of transcription of the gene, known as activated transcription, or repression of transcription, known as "silencing". Every gene has its own pattern of binding sites for transcriptional activators and repressors to ensure that the gene is only transcribed in the proper cell type(s) at the proper time during transcription. Therefore, studies on transcription factor proteins allows better understanding of these proteins and their roles in transcription, which ultimately aid in the studies of gene expression. The aim of our project was focusing on collect, clean and archive information about Transcription factors, their mode of action, their binding sites, domains, and genes they control. The Transcription factors and related information were collected from public databases such as PubMed, Central and Science Direct. It involved the following steps: a/ collection of binding sites of Transcription factors from the literature based on experimental evidence; b/ analyzing of protein domains of Transcription factors and matching these with the records in the above-mentioned and other databases; c/ analyzing links of Transcription factors and genes based on computational predictions; d/ analyzing links of Transcription factors and genes based on literature text-mining. A total of 59 Transcription factor binding sites (TFBS) from different literature articles were found and stored in a newly created database. The data collected will be used to build different tools/models for further used to build software related information. This data could also be used by researchers for designing TFBS related wet-lab experiments.

**Keywords:** Transcription factors, Computational predictions, database

**Inbix10/13*****In silico* prediction of T-cell epitopes from outer membrane protein of outer membrane vesicles of *Neisseria meningitidis* serogroup B (MC58) for vaccine design**

Sharat Chandra\* and Digvijay Singh

School of Biotechnology, Devi Ahilya University, Indore, India

\*Corresponding author: [sharat.chndr@gmail.com](mailto:sharat.chndr@gmail.com)**Abstract:**

*Neisseria meningitidis* serogroup B (MC58) is a leading cause of meningitis and septicaemia, principally infects the infants and adolescents. No vaccine is available for the prevention of these infections because the serogroup B capsular polysaccharide is unable to stimulate an immune response, due to its similarity with polysialic acid. To overcome this obstacles, we proposed to develop a peptide epitope based subunit vaccine from outer membrane (OM) protein contained in outer membrane vesicles (OMV) based on our in silico analysis. In OMV a total of 236 proteins were identified, only 15 (6.4%) of which were predicted to be located in outer membrane. The major requirement is the identification and selection of T-cell epitopes that act as a vaccine target. We have selected 13 out of 15 outer membrane proteins from OMV proteins. Due to similarity of the fkpA and omp85 with the human FK506 binding protein and SAMM50 protein have been removed from the analysis as their presence in the vaccine is likely to elicit an autoimmune response. ProPred and ProPred1 were used to predict promiscuous helper T Lymphocytes (HTL) and cytotoxic T Lymphocytes (CTL) epitopes in *N. meningitidis* serogroup B (MC58) respectively. Results indicate that porA, porB, opc, pilQ, nspA and mtrE are more suitable vaccine candidates. The predicted peptides are expected to be useful in the design of multi-epitope vaccines without compromising the human population coverage.

**Keywords:** Outer membrane vesicles, subunit vaccine, epitopes, *Neisseria meningitides* serogroup B (MC58).

**Acknowledgement:** We thank Dr. Tiratha Raj Singh for scientific support.

**Inbix10/14**

**Comparative analysis of computational epitope predictions: proposed library of putative vaccine candidates for HIV**

Dinesh Chaukiker, Arun Gupta\*

School of Computer Science & IT, DAVV, Indore  
Corresponding author: [dgupta.arun@gmail.com](mailto:dgupta.arun@gmail.com)

**Abstract:**

Designing a vaccine for a disease is one of the crucial tasks that involve millions and billions of dollars, several decades and yet there is no guarantee of successful results. Several pharmaceutical companies are investing their money and time in such activities. Computational biology could be of great help in these activities by providing a library of plausible candidates that might actually show some positive responses. Major Histocompatibility Complex (MHC) binding peptide prediction is one such area where the immense power of computers could be used to get a breakthrough. In this direction several databases and servers have been developed by many labs to predict the MHC binding peptides. These short peptides on the antigen surface are recognized by the MHC molecule and are presented to the receptors of T-cells for further immune response. Peptides that bind to a given MHC molecule share sequence similarity. Here we present a comparative study of servers that can predict the MHC binding peptides in a given protein sequence of the antigen. Based on this comparative analysis on HIV data, we are able to propose a library of putative vaccine candidates for the env GP-160 protein of HIV-1.

**Keywords:** MHC, MHC binding peptides, HIV-1, epitopic library, vaccine candidate

**Acknowledgment:** We would like to thank Dr. Tiratha Raj Singh, for the technical insights and support.

Note

