JPP JOURNAL OF PROTEINS AND PROTEOMICS
ISSN: 0975-8151: 8(3), 2017

Special Issue: "Bioinformatics and Computational Biology: Contributions from India"

Guest Editor

Prof. B. Jayaram
Department of Chemistry, Indian Institute of Technology, Delhi

CONTENTS

GENERAL INFORMATION, GUIDELINES AND POLICIES

The Journal: Journal of Proteins and Proteomics (JPP), administered by Proteomics Society, India (PSI), is a peer reviewed international journal envisaged to serve the world wide community of researchers and teachers dealing with the challenges of proteins and proteomics research resulting in an improved understanding of protein science in general. Published quarterly, the aim is also to supplement the regular issues with special issues annually in selected, relevant topics of protein science. The journal has an online presence at http://www.jpp.org.in. The journal publishes wide array of articles at no cost, whatsoever, to authors and provides free access to all articles through its website. Hard copies of the journal are available at nominal subscription charges.

Permissions: Please write to jppindia@gmail.com for information on how to request permissions to reproduce articles or any other information from the journal.

Disclaimer: The information and opinions presented in this journal reflect the views of the authors and not of the Proteomics Society or journal or its editors or international advisors or publisher and does not constitute endorsement by the journal or the society in any way. The journal or society does not assume any liability or responsibility for authenticity, correctness, accuracy, completeness or usefulness of the information published here and is the sole responsibility of the authors.

Plagiarism: The authors must ensure that they shun plagiarism in any form, whether in text material or data presented. Authors must thoroughly check their articles for plagiarism using standard, international tools and available practice and the journal assumes no responsibility for plagiarism committed by authors. Articles will be rejected or withdrawn if ever found guilty of plagiarism.

Ethical Issues, Rights: Authors are requested to conform to their institutional and country specific ethical guidelines and policies with respect to any biological sample. The society, journal or publisher carries no responsibility of any ethical mis-conduct. Proper ethical clearances must be obtained by the authors from appropriate authorities and the same must be declared in the published articles along with reference number and date of the clearance certificates. For human subjects and patient samples informed consent must be duly obtained by the authors as per regulations of the concerned authority and a statement to this effect should be included in the manuscript. Human and animal rights should not be violated and a statement to this effect must be included in the manuscript as well. All documents related to ethical issues must be readily available with the authors and must be produced on demand.

Conflict of interest: The authors must declare conflict of interests, if any.

Advertising Guidelines: JPP does accept classified advertisements from legal and well established agencies to promote the journal, as long as they conform to set policies of the journal and are related to the subject matter of the journal publications. Inquiries may be directed to jppindia@gmail.com. Advertisements do not however suggest that the journal endorses any of the products.

EDITORIAL NOTE: WELCOME TO THE TWENTY FIFTH ISSUE

The journal achieves yet another landmark in its sojourn – this issue is the 25th issue to be published since its inception. Over all these years, we have published on time and without any break whatsoever. To mark this achievement, this issue has been prepared as a special issue. This is the first time that we publish a special issue exclusively in the area of Bioinformatics and Computational Biology, which is an established area of intensive research in India in the current scenario. The country has seen rapid progress in Bioinformatics and Computational Biology over the last two decades and contributed significantly in terms of well cited publications, generation of data bases and tools, support to drug design and discovery programmes as well as dissemination of the knowledge among the younger generation and building up of infrastructure. In this light, Indian Institute of Technology, Delhi took the lead and organized a conference entitled "Breaking Barriers through Bioinformatics & Computational Biology" from 31st July to 1st August, 20017. The conference was organized by the Supercomputing Facility for Bioinformatics & Computational Biology, IIT Delhi under the auspices of Kusuma School of Biological Sciences, IIT Delhi (website: *http://www.scfbio-iitd.res.in*). Prof. B. Jayaram of IIT Delhi, one of the authorities in Bioinformatics and Computational Biology domain, led the event, which turned out to be an outright success. The conference boasted the presence of "Who's Who" in Bioinformatics and Computational Biology in India (as evident from the programme schedule) and was rich in its scientific content and outcome. The journal showcases this significant event in this special issue.

The Special Issue was Guest Edited by Professor B. Jayaram, who has highlighted the conference and other advances in India in the area in his Guest Editorial Column. The special issue carries four invited articles, all of which outlines the achievements in the field in different aspects. The abstracts of the conference are also included in the issue and present the entire landscape of bioinformatics and computational biology research that is ongoing in the country.

We take the opportunity to thank Prof. Jayaram for taking the lead. We thank the organizers of the conference for their vision and able support. We thank the sponsors who supported associated costs through advertisements that help the journal to be freely accessible and publish articles without any charge to authors. We thank PSI and the members of the editorial board for supporting and administering the journal. We thank the authors of the invited articles for their immense contribution. We indeed thank Mr. Shashank Shekhar, Supercomputing Facility for Bioinformatics & Computational Biology, IIT Delhi without whose valuable contributions this issue would not see the light of the day.

We hope you will appreciate our efforts and provide us with suggestions for improvement. Constructive criticism, suggestions and help of any kind, as always, are most welcome.

Suman Kundu
Editor-in-Chief

BREAKING BARRIERS THROUGH BIOINFORMATICS & COMPUTATIONAL BIOLOGY

Prof. B. Jayaram

Department of Chemistry, Supercomputing Facility for Bioinformatics & Computational Biology (SCFBio) & Kusuma School of Biological Sciences, Indian Institute of Technology, Delhi, Hauz Khas, New Delhi-110016, India
E-mail: bjayaram@chemistry.iitd.ac.in; bjayaram@scfbio-iitd.res.in; Website: www.scfbio-iitd.res.in

The big data in biology is creating an opportunity for both hypothesis and observation driven new science and new technologies. Celebrating the successes of Bioinformatics across the world and commemorating the 15th anniversary of the inauguration of the Supercomputing Facility for Bioinformatics and Computational Biology (SCFBio) at IIT Delhi, a national conference on "Breaking Barriers through Bioinformatics & Computational Biology", was held on July 31st and August 1st, 2017 at IIT Delhi. The conference brought together experts both young and senior, from diverse parts of the country to share their vision on emerging dimensions of Bioinformatics and Computational Biology while taking stock of the various softwares, tools and databases developed in the country. The conference witnessed lively presentations from 21 distinguished speakers and 60 young poster presenters showcasing their research work. About 200 people participated in the event. The lectures and posters covered scientific discoveries in progress in diverse laboratories across the country to overcome the burden of diseases among others. Bioinformatics software and visualization tools, it was noted, had been a key factor in the rapid and phenomenal advances in genomics, proteomics, medicine, drug discovery, systems approaches and in fact in every aspect of the new developments. The importance and necessity of various sequence, structural, functional and medical databases – generic as well as India specific – for new discoveries and for policy decisions was discussed. The conference also focused *inter alia* through panel discussion on how to leverage current strengths of the research community to seek answers to some fundamental questions in biology, to generate new technologies and employment opportunities and to identify what it takes India to become a leading contributor to the field of Bioinformatics and Computational Biology. The participants unanimously underscored the fact that India had a strong base in Bioinformatics. It was emphasized that bioinformatics driven translational research, with huge potential, needed stronger and wider collaborations, a mission mode approach and dissemination of the knowledge, softwares and products generated in the country through publications in high impact journals. Given the IT and quantitative skills and expertise in synthetic chemistry in India, it was projected that the country could do much more to benefit mankind.

About SCFBio. The Supercomputing Facility for Bioinformatics and Computational Biology (SCFBio), IIT Delhi (*www.scfbio-iitd.res.in*) was created with funding from the Department of Biotechnology (DBT), Government. of India, under the guidance of Principal Investigator, Prof. B. Jayaram with a vision to develop novel scientific methods, new softwares and to train the next generation of students and scientists in the area of Bioinformatics and Computational Biology. SCFBio was inaugurated on 31st July, 2002. It was converted into a Centre of Excellence (CoE) of BTIS network of DBT in December, 2013. SCFBio created

three main software suites namely *Chemgenome* (*http://www.scfbio-iitd.res.in/chemgenome/chemgenome3.jsp*) for genome annotation, *Bhageerath* (*http://www.scfbio-iitd.res.in/bhageerath/bhageerath_h.jsp*) for protein tertiary structure prediction and *Sanjeevini* (*http://www.scfbio-iitd.res.in/sanjeevini/sanjeevini.jsp*) for target directed lead molecule discovery and 43 web-servers for diverse bioinformatics applications. Recently, *Chemgenome* was able to fingerprint 7.2 million RNA (mRNA, tRNA, rRNA, miRNA etc.) genes encoded in 9282 prokaryotic and eukaryotic genomes based on physico-chemical properties (*Nucl. Acids Res.*, 2016; *DOI:10.1093/nar/gkw1236*) in a first of its kind. *Bhageerath*, ranked 11th by CASP organizers (http://predictioncenter.org) in the server category, is seen to be among the best web-servers globally in generating low-resolution models of protein tertiary structures (http://www.scfbio-iitd.res.in/Bhag+_CASP12.pdf). *Sanjeevini*, a complete drug design software suite, the only freely accessible server, has recently led to nanomolar antimalarials and a publication in the journal *Nature* (2016, DOI: 10.1038/nature16963). The genome to drug assembly line, *Dhanvantari* (*http://www.scfbio-iitd.res.in/software/dhanvantari_new/Home.html*), which combines the above three software suites is operational with possible entry at any point along the pipeline (Genome à Gene à Protein à Potential candidate drug molecules). The facility receives over 20,000 hits per day for its resources from users in more than 30 countries. The facility trained about 1000 students through its short-term and long-term training programmes in Bioinformatics and produced 15 PhDs, so far. Two start-up companies (Lead Invent & Novo Informatics) evolved out of the facility. A mobile app for *Sanjeevini* was released during the conference and made available at Google Play Store (*https://play.google.com/store/apps/details?id=com.sanjeevini*).

BJ and the entire SCFBio team are grateful to Prof. Suman Kundu and the editorial team of JPP as well as Proteomics Society, India, for readily accepting to publish the proceedings of the conference in the special issue of September, 2017.

Sponsors

Conference Proceeding

for

"BREAKING BARRIERS THROUGH BIOINFORMATICS & COMPUTATIONAL BIOLOGY"

31st July to 1st August, 2017

Venue : Seminar Hall, Main Building,
IIT Delhi Campus, New Delhi

Organizing Committee

Prof. B. Jayaram
Department of Chemistry, IIT Delhi

Prof. James Gomes
Kusuma School of Biological Sciences, IIT Delhi

Prof. Aditya Mittal
Kusuma School of Biological Sciences, IIT Delhi

Prof. D. Sundar
DBEB, IIT Delhi

Prof. Hemant Kashyap
Department of Chemistry, IIT Delhi

Shashank Shekhar
SCFBio, IIT Delhi

Technical Programme of the National Conference on"Breaking Barriers through Bioinformatics & Computational Biology July 31st and 1st Aug, 2017

Day 1 (Monday, July 31st, 2017)

SESSION - 1 (9:00 – 10:15): Inauguration **Chairperson: Prof. V. Ramgopal Rao, Director, IITD**	
09:00-09:15	Welcome by Prof. B. Jayaram, IITD
09:15-09:30	Dr. T. Madhan Mohan, Adviser, DBT, Guest of Honour
09:30-09:35	Release of Mobile App for *Sanjeevini* - a comprehensive indigenous drug design software suite - by Prof. Alok Bhattacharya, Chief Guest
09:35-09:50	Prof. Alok Bhattacharya, JNU; Chairman, Bioinformatics Task Force, DBTChairman, National Supercomputing Mission: Applications Group, Vice President, INSA.
09:50-10:05	Closing Remarks: Prof. V. Ramgopal Rao, Director, IITD
10:05-10:15	Vote of Thanks: Prof. Tapan K. Chaudhuri, IITD
High Tea (10:15-11:00)	
SESSION - 2 (11:00 – 13:00) **Chairperson: Prof. Alok Bhattacharya, JNU**	
11:00-11:30	Lecture 1: Prof. Indira Ghosh, JNU, New Delhi Bioinformatics driven translational research – Indian perspective
11:30-12:00	Lecture 2: Prof. Deepti Deobagkar, UoP, PuneSoftwares made in India
12:00-12:30	Lecture 3: Dr. Debasisa Mohanty, NII, New Delhi Databases made in India
12:30-13:00	Lecture 4: Dr. Suman Mallik, CDRI, LucknowBioinformatics publications from India
Lunch & Poster Presentation (13:00-14:30) in Lecture Hall Complex (3rd Floor, cafeteria) Chairpersons: Prof. Hemant Kashyap, IITD & Prof. Ashok Patel, IITD, Prof. T. Dutta, IITD & Dr. A. R. Rao, ICAR	
SESSION - 3 (14:30-16:30) Chairperson: Prof. James Gomes, IITD	
14:30-15:00	**Lecture 5: Prof. P. Satpati, IITG, Guwahati** Principle of Stop Codon Recognition by Eukaryotic Release Factor
15:00-15:30	**Lecture 6: Prof. R. Sankararamakrishnan, IITK, Kanpur** Anion-selective Formate/Nitrite Transporters: Bioinformatics analysis and molecular dynamics simulation studies
15:30-16:00	**Lecture 7: Prof. Sanjib Senapati, IITM., Chennai** Hydrated Ionic Liquids: A Novel Class of Solvents in Long-term DNA Stability
16:00-16:30	**Lecture 8: Mr. Sanjiv Shah (Intel)** Intel Solutions for Life Science
Coffee Break (16:30 - 16:45)	
SESSION - 4 (16:45- 17:15) Chairperson: Prof. N. Latha, SVC, DU, New Delhi About *Dhanvantari* (Genome to drug software suite developed at IITD) Genomics (*Chemgenome*) + Proteomics (*BhageerathH*) + Drug Design (*Sanjeevini*): SCFBio Team	
SESSION - 5 (17:15- 18:15) Chairperson: Prof. Aditya Mittal, IITD	
17:15-17:45	**Lecture 9: Prof. Manju Bansal, IISc., Bengaluru**Relative significance of genomic traits directs both expression level and breadth in plants
17:45-18:15	**Lecture 10: Prof. G. Reddy, IISc., Bengaluru** (Protein collapse and folding)

Breaking Barriers through Bioinformatics & Computational Biology July 31st and 1st Aug, 2017	
Day 2 (Tuesday, Aug 1st, 2017)	
SESSION - 6 (9:00 – 11:00) Chairperson: Prof. D. Sundar, IITD	
9:00 –9:30	**Lecture 11: Prof. B. Gopal, IISc., Bengaluru** A structural bioinformatics approach to understand gene expression in *Mycobacterium*
9:30 –10:00	**Lecture 12: Prof. Pinak Chakrabarti, Bose, Kolkata** A new protein secondary structure, *Topi* – implications in function and protein fibrillation
10:00 – 10:30	**Lecture 13: Dr. Binay Panda, GANIT Labs, Bengaluru** Data-driven view on human cancers: past, present and future**.**
10:30-11:00	**Lecture 14: Prof. Arnab Mukherjee, IISER., Pune** Partial Intercalation Kinks the DNA: Mechanism of Protein-DNA Intercalation
Tea Break (11:00-11:30)	
SESSION - 7 (11:30 – 13:30) Chairperson: Prof. Pradipta Bandyopadhyay, JNU, IITD	
11:30-12:00	**Lecture 15: Prof. G.P.S. Raghava, IIIT, New Delhi** InfoBioHealth (Informatics-based Biological Health) portal presents contribution of India for bioinformatics
12:00-12:30	**Lecture 16: Dr. Dinesh Gupta, ICGEB, New Delhi** Development of a computational method for prediction of protein arginine methylation
12:30-13:00	**Lecture 17: Prof. Prasad Bharatam., NIPER, Mohali** Design, Synthesis and Biological Evaluation of GSK-3b Inhibitors as Anti-Alzheimer's Agents
13:00–13:30	**Lecture 18: Dr. Uddhavesh Sonawane, CDAC, Pune** Structural insights into the chemically modified antisense /siRNA molecules using computational approaches
Lunch & Poster (13:30-14:30)	
Chairpersons: Prof. Manidipa Banerjee, IITD & Prof. Vivek Perumal, IITD Prof. Khushhali Menaria, MANIT, Bhopal & Prof. Archana Chugh, IITD	
SESSION - 8 (14:30-17:00) Chairpersons: Prof. P. Biswas, DU, Delhi & Prof. VibhaTandon, JNU, New Delhi	
14:30-15:00	**Lecture 19: Prof. P. Gautam, Anna Univ., Chennai** Modeling Proton Hopping in 8-hydroxyquinoline-5, 7 disulphonic acid
15:00-15:30	**Lecture 20: Prof. M. Gromiha, IITM., Chennai** Computational perspectives for understanding protein aggregation related human diseases
15:30-16:00	**Lecture 21: Prof. R. Bahadur; IITKgp, Kharagpur** Dissecting water binding sites at protein-protein interfaces
16:00-16:30	Lecture 22:Mr. Ashok Chaudhary (Fujitsu)
16:30-17:00	Lecture 23: : Mr. Asit Parija (MAPR)
Coffee Break (17:00-17:15)	
SESSION - 9 (17:15-18:00) Chairperson: Dr. Sharmila Mande, TCS, Pune **Computational Challenges in Microbiome Analytics **Panelists: Dr. Soma Marla, ICAR, New Delhi; Prof. Asad U Khan, AMU, Aligarh, Prof. B. Kundu, IITD	
17:15-18:00	Panel Discussion: Bioinformatics for new technologies and employment opportunities & how to leverage existing strengths to make India a leading contributor
SESSION - 10 (18:00-18:30) Chairperson: Dr. Gulshan Wadhwa, DBT, New Delhi & Prof. C. S. Dey, IITD & Prof. Shandar Ahmad, JNU, New Delhi (Judges for poster awards)	
18:00-18:30	Best 3 Poster presentations & Awards
18:30	Valediction

ABSTRACTS FOR SPEAKERS

L-1

Critical Analysis of publication lists from India in drug discovery process & impact of *in silico* designing tools/methods using Text based mining & association.

Pawan Kumar, Gourab Das & Indira Ghosh*

School of Computational & Integrative Sciences, Jawaharlal Nehru University, New Delhi 110067, E-mail: indira0654@gmail.com

Over the several decades, India is constantly challenged by communicable and non-communicable diseases which are originated either by poor life-style or by environmental factors. The pool of diseases is constantly posing serious threats to mankind especially among the poverty-stricken families [1]. Scientific communities from various parts of world are working hard to design drug molecules to overcome the burden of these life threaten diseases. In last three decades, many computational algorithms and tools have been developed to identify potential drug targets and their inhibitors [2][3]. It is believed that computational techniques have reduced the time and money required to develop an inhibitor into drug [4]. However, applicability and deliverability of these *in silico* techniques in rational drug designing is not fully evaluated. In the present study, a PubMed/Medline extracted (using Title, Author, Affiliation, Abstract & Keywords only) data driven analysis has been performed to highlight the influence and progress of the theoretical methods in the field of medicinal research/drug discovery, across India and compared with the World. Published in Literature and manually curetted drug discovery related keyword dictionaries have been built and utilized to filter these abstracts from different branches of life sciences. Point wise mutual information (PMI) [5] has been used for association analysis and its impact on the final inhibitor/drug designing. Most important observations are that drug discovery has been an interdisciplinary research progressively used many tools starting with QSAR, docking, pharmacophore, Molecular Simulations etc. The publications contributed from India (2%) is similar in compare to the contribution in total world publications, suggesting large scope in future. Data coverage as represented since 1990-2015 in PubMed as indicated by number of publications associated in drug discovery is almost same in world and India (~80%). Emerging institutes/Universities are contributing since last 10 years as observed from Indian publication list. However, this method has many limitations as discussed.

Reference

[1] T. Dikid, S. K. Jain, A. Sharma, A. Kumar, and J. P. Narain, "Emerging & re-emerging infections in India: An overview," *Indian Journal of Medical Research*. 2013.

[2] P. M. Njogu, E. M. Guantai, E. Pavadai, and K. Chibale, "Computer-Aided Drug Discovery Approaches against the Tropical Infectious Diseases Malaria, Tuberculosis, Trypanosomiasis, and Leishmaniasis," *ACS Infect. Dis.*, vol. 2, no. 1, pp. 8–31, 2016.

[3] T. Katsila, G. A. Spyroulias, G. P. Patrinos, and M.-T. Matsoukas, "Computational approaches in target identification and drug discovery," *Comput. Struct. Biotechnol. J.*, 2016.

[4] I. M. Kapetanovic, "Computer-aided drug discovery and development (CADDD): In silico-chemico-biological approach," *Chem. Biol. Interact.*, 2008.

[5] G. Bouma, "Normalized (Pointwise) Mutual Information in Collocation Extraction," *Proc. Ger. Soc. Comput. Linguist. (GSCL 2009)*, pp. 31–40, 2009.

L-2

Bioinformatics Software from India: Challenges and Promises

Deepti Deobagkar

ISRO Chair Professor, SavitribaiPhule Pune University, Pune, 411007, E-mail: deepti.deobagkar@gmail.com

Bioinformatics software and visualisation tools have been a key factor in the rapid and phenomenal advances in genomics, proteomics, medicine, drug discovery, systems approaches and in fact in every aspect of new developments. The contributions from India related to Bioinformatics software which have been important in a few specific areas will be reviewed. India has a strong hold in computation and IT and has a pool of bright and young talent with demographic dividend along with experienced and excellent mentors and researchers. It also has an advantage of an early start and extensive and

organised network in the Bioinformatics education and research with substantial inputs from the Indian government. Although small in number and scale, the Bioinformatics Industry is also making its mark felt in India.

It is important to consolidate and take bold and visionary steps which will position us to take up the challenges by forming and consolidating networks of the strong IT and computational industries as well as training and research institutes along with state of art biology and computational biology researchers to address the emerging and future challenges in the areas of drug discovery, genomics, metagenomics, protein engineering, systems and network approaches, neurobiology and cognition, agriculture, robotics and electronics among others. With the large pool of human resource, it will be important to strategically utilise the skill development with strong academic input with planning and vision to drive the engine of software development to make original and path breaking contributions. This can be done by working in a mission mode with time targeted goals in small, focussed multi and interdisciplinary groups which will work by building strategies with a problem-solving approach. It will be crucial to address the simple problems but not shy away from taking up the tough challenges thrown in by the massive explosion in the genomic and proteomics data and in unravelling the complexities of life. A new approach and fresh outlook for the way forward is critical in order to meet the challenges and fulfil the promises.

L-3

Databases for biological sciences developed in India

Debasisa Mohanty

Bioinformatics Center, National Institute of Immunology, Aruna Asaf Ali Marg, New Delhi – 67, India
E-mail: deb@nii.res.in, deb@nii.ac.in

The complexity of biological systems requires use of a variety of experimental methods with ever increasing sophistication to probe various cellular processes at molecular and atomic resolution. The availability of technologies for determining nucleic acid sequences of genes and atomic resolution structures of biomolecules prompted development of major biological databases like GenBank and PDB almost four decades ago. India was one of the few countries to realize early, the utility of such databases for progress in modern biology/ biotechnology. DBT, India established Biotechnology Information System (BTIS) network in late eighties. Starting with the genome sequencing revolution at the turn of the century, application of high-throughput sequencing technologies in biology and medicine for analysis of genomes, transcriptomes, epigenomes and microbiomes have generated massive volumes of sequence data. BTIS network has not only provided state of the art computational infrastructure to research institutes and universities for utilizing various biological databases developed abroad in their research, it has also actively promoted R&D projects in Bioinformatics to develop a variety of biological databases in diverse areas. It is encouraging to note that, a large number of biological databases or data driven software tools developed in India, have been published in leading peer reviewed international journals like Nucleic Acids Research, Bioinformatics, Database, BMC, PLoS and NPG series publication. Some of these databases are not only unique, they are also highly accessed as reflected in number of citations. Apart from databases developed by individual research groups, BTIS has initiated consortium projects to develop major India centric databases on Mycobacterium tuberculosis, Rice and Mango, which can potentially have practical applications in health and agricultural sciences. Many of these biological databases has also helped in development of novel data mining methods, prediction strategies and data driven application softwares or web servers. The talk would given an overview of databases developed in India, their impact on data driven research in biology and also India centric future plans for making transitions to big data revolution in biology by combining techniques like Deep Learning with biological big data.

L-4

Rise of Bioinformatics & Computational Biology in India: A look through publications

Suman K Mallik

Chief Scientist & Coordinator, BTIS
CSIR-Central Drug Research Institute, Lucknow
E-mail: sumanmallik@cdri.res.in

Computational biology and bioinformatics have been part and partial of biomedical research for few

decades now. However, the institutionalisation of bioinformatics research took place with the establishment of Distributed Information Centres (DISCs) in the research institutions of repute in various disciplines by the Department of Biotechnology, Govt. of India. Though, at initial stages, this endeavour was mainly focussed on providing infrastructure for using information technology and internet based communication and tools for carrying out computational biology and *in-silico* assisted research in varied arena of research starting from disease biology to agricultural crops, spices, veterinary science and many more, the natural outcome of establishment of such facilities resulted into new experiments with bioinformatics tools. This BTIS grew into a solid movement and a large number of publications started coming out of these centres. In the end of last century, bioinformatics started developing like a full-fledged research subject.

In the last decade, a need was felt to actually make a factual estimation of the result of this endeavour of DBT which had, by then, established over a hundred centres in almost all disciplines of biomedical research.

In a bid evaluate the efforts and outcome of these centres, we were entrusted with collecting and collating the publications of these centres. However, when the full data was compiled, the DBT task force felt that the study must include Non-BTIS Centres also so as to expand the report to have a glimpse of bioinformatics publications from the country.

The compiled result was quite encouraging and a compendium was published by DBT. This compendium was not only appreciated but the process of collecting information continued hereafter. DBT encouraged us to have annual compilation of the centres and even started annual incentive awards for best papers in bioinformatics and computational biology.

The present presentation is all about the result of all these efforts of collection and compilation. The data on publications will be presented such as number of publications, subject-wise papers and results based on various parameters. The actual problem in collecting such data will also be elaborated giving same glimpses of how they could be dealt with. Overall, this presentation is an effort to look at the growth of bioinformatics in India through the publications in reputed journals by a large number of research organisations.

L-5

Principle of Stop Codon Recognition by Eukaryotic Release Factor

P. Satpati

IITG, Guwahati

In termination of protein synthesis, eukaryotic release factor (eRF1) binds to mRNA stop codons (UAA, UAG or UGA) at ribosomal A-site and catalytically hydrolyzes the nascent polypeptide chain from P-site tRNA. Recently determined cryo-EM structures of termination complexes (eRF1 bound stop codon programmed ribosome) have provided atomic insight into the interaction pattern. However the energetic principle for stop codon decoding by eRF1 is a fundamentally unsolved problem. Using the cryo-EM structures as templates, we carried out molecular dynamics simulations of stop and sense codon programmed termination complexes. The focus of my talk will be to address the following questions based on our structure based molecular dynamics simulations:

(a) How strongly sense codons are discriminated with respect to sense codons by eRF1?

(b) How the strength of discrimination is linked to the 3D structures of the sense and stop codon programmed termination complexes?

L-6

Anion-selective Formate/Nitrite Transporters: Bioinformatics analysis and molecular dynamics simulation stuides

R. Sankararamakrishnan

Department of Biological Sciences and Bioengineering Indian Institute of Technology Kanpur, Kanpur 208016, India, E-mail: rsankar@iitk.ac.in

Channels and transporters are integral membrane proteins that selectively transport ions and other solutes across the membrane. The family of Formate/Nitrite transporters selectively transports monovalent anions such as formate, nitrite and hydrosulphide which are main metabolites of

bacterial respiration during anaerobic mixed-acid fermentation [1]. When accumulated in the cytoplasm, these anions become cytotoxic. Individual members that selectively transport formate, nitrite and hydrosulphide have been investigated experimentally. Three-dimensional structures of FNTs indicate that they share the same hourglass helical fold with aquaporins and aquaglyceroporins [2, 3]. As in aquaporins, FNTs also have two constriction regions, namely, cytoplasmic slit and central constriction. Members of FNTs are found in bacteria, archaea, fungi and protists. However, no FNT homolog has been identified in mammals. With FNTs as potential drug targets for many bacterial diseases, it is important to understand the family, its diversity, taxonomic distribution, evolution and the molecular mechanism of transport.

In this talk, I will first discuss our bioinformatics analyses of 2206 FNT sequences from bacteria, archaea and eukaryotes. FNT sequences are very both sequentially and functionally diverse. However, homology modeling followed by structure-based sequence alignment revealed that nearly one third of all the positions within the transmembrane region exhibit high conservation across all three kingdoms. Phylogenetic analysis of prokaryotic FNT sequences revealed eight different subgroups. Formate, nitrite and hydrosulphide transporters respectively are clustered into two (FocA and FdhC), three (NirC-α, NirC-β and NirC-γ) and one (HSC) subfamilies. We have also recognized two FNT subgroups (YfdC-α and YfdC-â) with unassigned function [4]. Analysis of taxonomic distribution indicates that each subfamily prefers specific taxonomic groups. Certain positions in the two constriction regions and some interior-facing residues display subfamily-specific conservation. We have developed dbFNT, a database of FNT models and associated details. dbFNT (*http://bioinfo.iitk.ac.in/dbFNT*) is freely available to scientific community. In the second part of the talk, I will describe our simulation studies of representative members of FNT family in explicit bilayers. We have investigated the transport mechanism of formate in both ionized and in neutral form and the results will be discussed.

References

[1] S. Leonhartsberger, I. Korsa and A. Bock, *J. Mol. Microbi. Biotech.* 4, 269-276 (2002).

[2] R. K. Verma, A. B. Gupta and R. Sankararamakrishnan, *Methods Enzymol.* 557, 485-520 (2015).

[3] A. B. Waight, J. Love, D.-N. Wang, *Nature Struct. Mol. Biol.* 17, 31-37 (2010).

[4] M. Mukherjee, M. Vajpai and R. Sankararamakrishnan, *BMC Genomics* (2017) in press.

L-7

Hydrated Ionic Liquids: A Novel Class of Solvents in Long-term DNA Stability

Sanjib Senapati

Dept of Biotechnology, I.I.T. Madras, Chennai 600036, India, E-mail: sanjibs@iitm.ac.in

Very recently, hydrated ionic liquids (ILs) have been identified as ideal media for various biological applications, including the DNA storage at room temperature. Hence, understanding the binding characteristics and molecular mechanism of interactions of ILs with DNA is of both practical and fundamental interest. We employ molecular dynamics (MD) simulations and spectroscopic experiments to unravel the key factors that stabilize DNA in hydrated ILs. Both simulation and experimental results show that DNA maintains the native B-conformation in ILs. Simulation results further suggest that, apart from the electrostatic association of IL cations to DNA backbone, groove binding of IL cations through hydrophobic and polar interactions contribute significantly to DNA stability. CD spectral measurements and fluorescent dye displacement assay confirm the intrusion of IL molecules into the DNA minor groove. In the later half of the talk, I will briefly discuss how ILs can maintain dehydrated DNA in B-form.

References

[1] C. Aneesh, D. Ghoshdastidar, S. Senapati, Groove binding mechanism of Ionic liquids: A key factor in long-term stability of DNA in hydrated Ionic liquids? *J. Am. Chem. Soc.,* 134, 20330, (2012).

[2] D. Ghoshdastidar, S. Senapati, Ion-Water Wires in Imidazolium-Based Ionic Liquid/Water Solutions Induce Unique Trends in Density. *Soft Matter*, 12, 3032-3045 (2016).

[3] D. Ghoshdastidar, D. Ghosh, S. Senapati, High Nucleobase-Solubilizing Ability of Low-Viscous Ionic Liquid/Water Mixtures: Measurements and Mechanism. *J. Phys. Chem. B*. 120, 492-503 (2016);

L-8

Intel Solutions for Life Science

Mr. Sanjiv Shah

Vice president of the Software and Services Group and General manager of Technical, Enterprise and Cloud Computing software tools at Intel Corporation

Abstract: Intel® Corporation is not commonly associated with life science or biomedicine. However, Intel has played a role in the computational life sciences for decades, and more recently has been at the forefront of genomic medicine. This talk will look at the computational barriers to widespread implementation of precision medicine, and describe how Intel hardware, software, and enabling are helping to remove these barriers. The talk will conclude with a discussion of the Intel® Scalable System Framework, its role in democratizing large-scale computing, and its applicability to future use-cases in computational life science and genomic medicine that combine traditional HPC, big data analytics, and artificial intelligence.

L-8a

Productivity and High-Performance for Computational Life Science

Mr. Sanjiv Shah

Vice president of the Software and Services Group and General manager of Technical, Enterprise and Cloud Computing software tools at Intel Corporation

Abstract: The computational landscape in life science has historically been heavily fragmented and dominated by academic software development. Life sciences applications come and go with much higher frequency than other scientific disciplines. Consequently, there are only a few dominant applications (e.g., BLAST, AMBER, BOWTIE) at any given moment that warrant extensive architectural tuning. Such a rapidly shifting application landscape requires a rapid prototyping and deployment cycle. This is where productivity languages like Python really shine. The Intel® Distribution for Python combined with the Intel® performance libraries gives life science application developers both productivity and performance on modern processors.

L-9

Relative significance of genomic traits directs both expression level and breadth in plants

Manju Bansal and Sanjukta Das

Molecular Biophysics Unit, Indian Institute of Science Bangalore

Understanding the structural organisation of genome and various gene properties has helped in elucidating gene regulation and expression at cellular and molecular levels. In higher eukaryotes, gene structure and promoter architecture have emerged as significant factors influencing variation in number of transcripts (expression level) and specificity of gene expression in a tissue (expression breadth), which eventually shape the phenotype. In this study, transcriptome data of different tissue types at various developmental stages of *A. thaliana*, *O. sativa*, *S. bicolor* and *Z. mays* have been used to understand the relationship between properties of gene components and its expression. Our findings revealed that in plants gene architecture as well as promoter properties are significantly linked to gene expression level as well as breadth, suggesting interconnected molecular and functional pathways. Interestingly, results from multiple regression analysis, which predicts a significant amount of change in expression level and breadth is influenced by a particular genomic trait, revealed that intron content of primary transcript (as %) is a powerful determinant of tissue specificity. Similarly, among structural properties of the promoter, stability was found to be negatively linked to expression breadth, while DNase1 sensitivity strongly governed gene expression breadth in monocots and gene expression level in dicots. This differential response to DNase1 sensitivity exhibited by different plants led us to examine the compositional distinction in their promoter regions, and further analysis revealed that tissue specific genes are enriched with TATA box and Y-patch. An extensive study on the size of the orthologous groups and gene expression parameters illustrated that multi-copy orthologous genes in plants are longer, highly regulated and tissue specific.

L-10

Protein collapse and folding

Govardhan Reddy

IISc, Bengaluru

A fundamental question in protein folding is whether the coil to globule collapse transition occurs during the initial stages of folding (burst-phase) or simultaneously with the protein folding transition. Single molecule fluorescence resonance energy transfer (FRET) and small angle X-ray scattering (SAXS) experiments disagree on whether Protein L collapse transition occurs during the burst-phase of folding. We study Protein L folding using coarse-grained model and molecular dynamics simulations. The collapse transition in Protein L is found to be concomitant with the folding transition. In the burst-phase of folding, we find that FRET experiments overestimate radius of gyration (Rg) of the protein due to the application of Gaussian polymer chain end-to-end distribution to extract Rg from the FRET efficiency. The actual decrease in Rg is close to the statistical uncertainties of the Rg data measured from SAXS experiments, which suggest no compaction, leading to a disagreement with the FRET experiments.

L-11

A structural bioinformatics approach to understand gene expression in *Mycobacterium tuberculosis*

B. Gopal

Indian Institute of Science (IISc) Bengaluru

A multilayered regulatory mechanism governs gene expression in *Mycobacterium tuberculosis*. This ensures an appropriate cellular response to diverse micro-environments encountered in the host. Computational strategies to integrate biophysical, biochemical and structural information are important to understand this regulatory network. An area of interest is the cross-talk between apparently different networks and the role of leaderless mRNA. In this presentation, we describe some of these approaches and the challenges to be addressed in this area.

L-12

A new protein secondary structure, *Topi* – implications in function and protein fibrillation

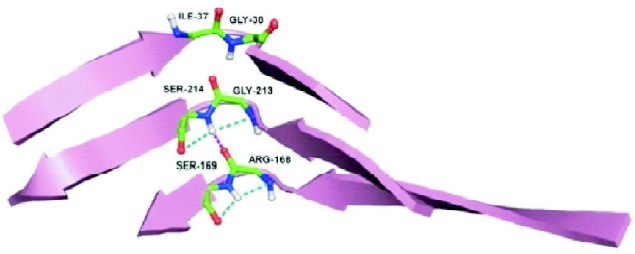
Pinak Chakrabarti

Bioinformatics Centre, and Department of Biochemistry, Bose Institute, P1/12 CIT Scheme VIIM, Kolkata 700054
E-mail: pinak@jcbose.ac.in

Protein chain is arranged into regions in which the backbone is organized into regular patterns (of conformation and hydrogen bonding) to form the most common secondary structures, α-helix and β-sheet. A new secondary structure, *topi*, is discussed in which a pair of 2-residue segments, each containing five-membered fused-rings motifs, distant in sequence are linked by hydrogen bond (Figure) [1]. Though a small motif, it appears to be important in the context of local folding patterns of proteins and occurs near protein active site. The motif shows quite significant residue preference, and a Cys (or Ser) occupying the second position may further stabilize the motif by forming additional hydrogen bond across it. Remarkably, *topi* is found within disease causing misfolded proteins, such as fibrilled form Aâ42, and within the interface region between the chains of some multimeric proteins. The motif may be an important component of fibrillation and useful for modeling loop regions.

Reference

[1] Dhar J., Kishore, R. and Chakrabarti P. A novel secondary structure based on fused five-membered rings motif. *Sci. Rep.* **6**, 31483; doi: 10.1038/srep31483 (2016).



L-13

Data-driven view on human cancers: past, present and future

Binay Panda

GANIT Labs, Bengaluru

The talk will address the following:

- Why cancer needs to be re-defined?
- What role large amount of data will play in our understanding of cancer and for its diagnosis, prognosis and therapy choices?
- Data integration towards clinical solutions in head and neck cancer
- The "jargon" Big data
- How big is big data in biology?
- Future of cloud computing in academic research
- Why sharing data freely and pre-publication is a sane and sensible thing?

L-14

Partial Intercalation Kinks the DNA: Mechanism of Protein-DNA Intercalation

Arnab Mukherjee

IISER, Pune

In this talk, we will discuss the origin of kink formation in DNA observed for some transcription factor (TF) protein-DNA complexes. We investigate the molecular origin of the DNA kinks using small molecule intercalation pathway, crystallographic analysis, and free energy calculations involving four different transcription factor (TF) proteins and their complexes with DNA. We find that although protein binding may bend the DNA, it alone is not sufficient to kink it. We will show that partial, not complete, intercalation is required to form kink at a particular place in DNA. We find that while amino acid alone can induce the desired kink through partial intercalation, protein provides thermodynamic stabilization of the kinkedstate in TF-DNA complexes. Further, we will discuss the overall mechanism of protein intercalation mechanism into DNA.

L-15

InfoBioHealth (Informatics-based Biological Health) portal presents contribution of India for bioinformatics

G. P. S. Raghava

IIIT, New Delhi; Web Site: http://webs.iiitd.edu.in

In this lecture, I will describe major features of our newly developed web portal on health informatics. This web portal maintained wide range of servers, databases and software developed in the field of bioinformatics, chemiinformatics, immunoinformatics, clinical bioinformatics, health informatics, genomics, etc. The main purpose of this web protal is to provide help to biologist working in the field of vaccine development, drug desigining, etc. Overall aim of the web portal is to provide scientific computation and resources required in the healthcare sector. There are number of scientific fields where India's contribution is enormous, one of the field is computational biology or bioinformatics. One of the fine contribution of India in 1960's in the field of computational biology is G.N. Ramachandran Plot. Department of biotechnology, systematically initiate BTISNET (Biotechnology Information System Network) in 1987, which have more than 200 bioinformatics centres. There is a tremendous change in the filed of bioinformatics in India, traditionally computer programs are developed for their own use.

In last two decades, number of groups emerged in India, who are developing open source software. These groups are providing web-based scientific services to research community. In past number of attempts have been made to present Indian contribution in the field of bioinformatics. In this talk, I will talk about InfoBioHealth (Informatics-based Biological Health), which is under developent at my group. Major aim of this portal is to compile web-based services in the filed of bioinformatics particularly in the field of biomedical sciences. This portal will collect and compile web-based services maintained by Indian institutes or research groups. This will be useful to understand contrbution of India in the field of computational biology.

L-16

Development of a computational method for prediction of protein arginine methylation

Dinesh Gupta

Group Leader, Translational Bioinformatics Group, International Centre for Genetic Engineering and Biotechnology (ICGEB), Aruna Asaf Ali Marg, New Delhi, 110 067

Protein Post-Translational Modification (PTMs) are important amino acid modifications which carry out and regulate several important biological functions, including gene regulation and signal transduction. Experimental identification of arginine methylation site is arduous task, costly, as well as time and labour-intensive too. Computer Aided Prediction tools for arginine methylation play an important task in rapid screening, identification of possible methylation sites in proteomes to aid experimental studies. However, the available prediction methods are not so efficient and yield false positive results. In order to develop a more efficient prediction algorithm for arginine methylation, we conducted a fresh comprehensive sequence data collection, analysis, and selection of relevant sequence based features, that led to development of a simple machine learning based prediction tool. The recently developed method by us performs reasonably better as compared to the existing tools.

L-17

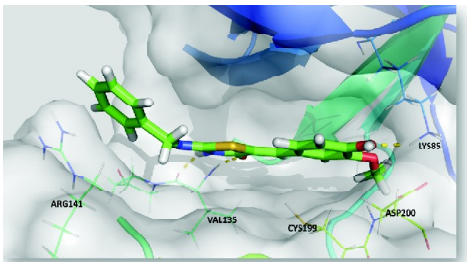
Design, Synthesis and Biological Evaluation of GSK-3b Inhibitors as Anti-Alzheimer's Agents

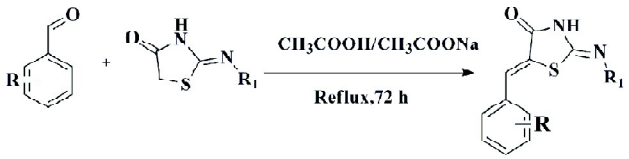
Prasad V. Bharatam

[a]*Department of Medicinal Chemistry, National Institute of Pharmaceutical Education and Research (NIPER), S.A.S. Nagar (Mohali), 160062 Punjab, India*
**E-mail:pvbharatam@niper.ac.in*

GSK-3β is an important enzyme implicated in diabetes, Alzheimer's disease, cancer and malaria. Inhibition of this enzyme can potentially solve some of the above ailments. Pharmacoinformatics methods can be successfully employed to design GSK-3b inhibitors. Molecular docking methods provide information regarding the interaction between ligands and the enzyme. QSAR methods are useful in designing new leads. Pharmacophore mapping methods are useful to perform virtual screening of compounds to identify new leads. Many other pharmacoinformatics methods can be used to perform lead optimization. Especially, molecular dynamics methods can be used to describe the stability of the inhibitor-enzyme interactions. Quantum chemical methods can be used to know the electronic details of the leads.

5-benzylidene-2-iminothiazolidin-4-one derivatives were designed using pharmacoinformatics techniques. Selectivity of these ligands was examined against CDK-2 and CDK-5. The identified compounds were then taken up for synthesis, followed by biological evaluation. The biological evaluation resulted in the identification of ten compounds to be active in lower nano molar range (2.1 to 85.4nM), in "fluorescence resonance energy transfer" biochemical assay. Out of thesethree have been found to selective against CDK-2.





References

[1] N. Tripathi et al, **O**ncotarget, 2016. In press.

[2] P.V. Bharatam, C.N. Kundu, Scientific Reports (Nature), 2016, 10.1038/srep20600.

[3] P.V. Bharatam, *et al*., Chem. Eur J., 2016, 22, 1088-1096.

[4] G.K. Rastogi, *et al*., ACS Med. Chem. Letts., 2015, 1065-1070.

[5] A. Bhaviskar *et al*., ACS Med. Chem. Lett., 2015, 6, 481-485.

[6] A. Baviskar, *et al*, J. Med. Chem., 2011, 54, 5013-5030.

[7] P. V. Bharatam, D.S. Patel, P. Iqbal., J. Med. Chem., 2005, 48, 7615-7622.

[8] S. Khanna, *et al*, J. Med. Chem., 2005, 48, 3015-3025.

L-18

Structural insights into the chemically modified antisense /siRNA molecules using computational approaches

Mallikarjuna Chari UVN and Uddhavesh B. Sonavane*

High Performance Computing - Medical & Bioinformatics Group, Centre for Development of Advanced Computing (C-DAC), Pune University Campus, Pune-411007, India
**E-mail: uddhaveshs@cdac.in*

The studies on the antisense-based therapy and the RNAi based therapy are developing hope for the treatment of various critical diseases. In antisense strategy, the design of DNA based modified antisense oligomer is required to bind to the target mRNA on which RNAase H binds and degrades target mRNA molecule. In case of RNAi strategy, the design of RNA based modifications in guide strand is needed to bind to passenger strand or target strand mRNA on which human AGO2 binds and cleaves mRNA only. In both the methodologies, the detailed experimental and computational study of chemically modified nucleotides much helps in understanding of the mechanism of action. The computational studies of these chemically modified nucleotides at different levels like monomer level, oligomer duplex level and oligomer duplex binding to the respective enzyme definitely throw more light in structural aspects, which are crucial in understanding and designing novel modifications and identifying the position of modification to be incorporated for best effect. We have studies few important antisense modifications (PS, MOE, RcMOE, ScMOE, LNA) at monomer level using quantum calculations. Molecular dynamics simulations of antisense oligomer-RNA duplexes (antisense oligomer in gapmer form base-paired with complementary RNA oligomer) have been carried out. The results of multiple trajectories of each duplex provide insights about the binding energies, solvent accessible surface area and other structural parameters. The free energy binding values are showing similar trend as shown in experiments like LNA contained ASO is strongly binding to complementary RNA. The solvent accessible surface area of the duplexes shows that the MOE contained ASO have highest values compared to other modified ASOs, which may be influencing the toxicity. This may be the reason behind in selecting MOE based ASO oligomer for targeting mRNA of ApoB and released as a drug (Mipomersen) into the market for the treatment of hypercholesteramia by Ionis Pharmaceuticals.

L-19

Modeling Proton Hopping in 8-hydroxyquin oline-5, 7 disulphonic acid

P. Gautam

Centre for Biotechnology & AU{KBC Research Centre , Anna University, Chennai 600025}
sites.google.com/site/gautampena

The hydrated proton H^+ plays a fundamental role in chemical and bio- chemical processes occurring in both homogeneous and inhomogeneous aqueous environments, such as proton transport in bulk water and proton pumping through membrane proteins. The latter is an archetypical example for processes occurring in biological environments, where the proton is solvated and transferred in partially aqueous environments especially in reactive sites of enzymes. These biological environments are poorly represented by model systems in bulk water. A variety of compounds have been synthesized in low water environ- ments to understand and model the mechanism of proton hopping. Two main intermediates are seen during this process, the Zundel ion and the Eigen Ion. We have synthesized 8-hydroxyquinoline-5, 7 disulphonic and recrystallized it in methanol to strip away molecules of water. The structure of the molecule revealed that Zundel ion was stabilized in the crystal. Ab- initio molecular dynamics simulation was then carried out to understand the dynamics of proton hopping in this complex. During the course of simulation, the Zundel ion coordinates with a water molecule to form an open $H_7O_3^+$ structure. This transition state structure de-solvated rapidly forming Zundel ion facilitating proton hopping in the first solvation shell. One of the sulphonic acid groups in the 5 or 7 position of the 8-hydroxyquinoline 5,7 disulphonic acid bonds with the Zundel ion favoring the proton to be transferred to the nearby water molecule through the formation of proton defects. The simulation results support the structural di usion mechanism

and that charged complex migrates through the hydrogen bond network.

L-20

Computational approaches for understanding protein aggregation related human diseases

M. Michael Gromiha

Department of Biotechnology, Indian Institute of Technology Madras, Chennai 600036

Aggregation is a process, which prevents proper folding of proteins and it is necessary for all organisms to overcome aggregation for maintaining their native states. The aggregation of proteins causes several neurodegenerative human diseases including Parkinson disease and Alzheimer disease. Given the importance of amyloid fibril formation in different areas of biology, it is important to elucidate the mechanisms for aggregation as well as to identify the probability of peptides to form aggregation along with aggregation related diseases. We have addressed these problems on different perspectives: (i) development of a curated protein aggregation database (1), (ii) development of position specific and residue pair potentials using amyloid-forming hexapeptides and non-amyloids (2) (iii) discrimination of amyloid forming peptides and non-amyloids using various features of amino acid residues, and positions specific residue and residue-pair potentials (3), and (iv) coupling with immunogenicity to address autoimmune diseases (4). Recently, we have scanned experimentally known aggregating peptides with human genome and revealed the presence of aggregation prone regions in proteins, which are involved in diseases (5). The salient features of the results will be discussed.

References

[1] A.M. Thangakani, R. Nagarajan, R. Sakthivel, S. Kumar. D. Velmurugan and M.M. Gromiha (2016) PLoS One. 11:e01529493.

[2] A.M. Thangakani, S. Kumar, D. Velmurugan and M.M. Gromiha (2012) BMC Bioinformatics 14(Suppl 8), S6.

[3] A.M. Thangakani, S. Kumar, R. Nagarajan, D. Velmurugan and M.M. Gromiha (2014) Bioinformatics, 30, 1983-1990.

[4] S. Kumar, A.M. Thangakani, R. Nagarajan, S.K. Singh, D. Velmurugan and M.M. Gromiha (2016) Sci Rep. 6:22258.

[5] R. Prabakaran, D. Goel, S. Kumar, M.M. Gromiha (2017) Proteins 85:1099-1118.

L-21

Dissecting water binding sites at protein-protein interfaces

Sunandan Mukherjee, Chandran Nithin and Ranjit Prasad Bahadur*

Computational Structural Biology Lab, Department of Biotechnology, Indian Institute of Technology Kharagpur, 721302, India

**E-mail: r.bahadur@hijli.iitkgp.ernet.in*

We dissect the protein-protein interfaces into water preservation (WP), water hydration (WH) and water dehydration (WD) sites by comparing the water mediated hydrogen bonds (H-bond) in the bound and unbound states of the interacting subunits.Upon subunit complexation, if a H–bond between an interface water and a protein polar group is retained, we assign it as WP site; if it is lost, we assign it as WD site and if a new H–bond is created, we assign it as WH site. We find that the density of WD sites is highest followed by WH and WP sites except in antigen and (or) antibody complexes, where the density of WH sites is highest followed by WD and WP sites. Furthermore, we find that WP sites are the most conserved followed by WD and WH sites in all class of complexes except in antigen and (or) antibody complexes, where WD sites are the most conserved followed by WH and WP sites. A significant number of WP and WH sites are involved in water bridges that stabilize the subunit interactions. At WH sites, the residues involved in water bridges are significantly better conserved than the other residues. However, no such difference is observed at WP sites. Interestingly, WD sites are generally replaced with direct H-bonds upon subunit complexation. Significantly, we observe many water mediated H-bonds remain preserved in spite of large conformational changes upon subunit complexation. These findings have implications in predicting and engineering water binding sites at protein-protein interfaces.

ABSTRACTS FOR POSTER

P-1

Sanjeevini mobile application for Android devices

Abhilash Jayaraj[1,2*], Mano Teja Boyapati[3], Durgesh Choudhary[3] , Shashank Shekhar[2] and B. Jayaram[1,2,4*]

[1]Department of Chemistry, [2]Supercomputing Facility for Bioinformatics & Computational Biology, [3]Department of Computer Science, [4]Kusma school of Biological Scineces, Indian Institute of Technology, Delhi, Hauz Khas, New Delhi-110016, India
*E-mail: [*]abhilash@scfbio-iitd.res.in; *bjayaram@chemistry.iitd.ac.in*

In today's fast paced world, drug discovery is still a costly and time consuming process. Some of this time and cost can be cut down using Computer Aided Drug Discovery (CADD). CADD requires application of streamlined pipeline of programs and methodologies to predict a hit molecule with theuraputic value against a protein/nucleic acid target. The concerted efforts of both *in silico* and wet lab studies is expected to yield candidate drug molecules.

Commercial *in sillico* drug discovery solutions addressing this problem tend to be costly and not accessible to all academic labs. To this end we have worked towards converting our *Sanjeevini*[1] methodology for drug discovery, into a freely accessible android mobile application.The application includes a variety of drug discovery modules for Active site prediction, ligand screening, docking and scoring against both protein and DNA targets. It also provides access to a million molecule database which can be assessed based on physico-chemical properties to identify hit molecules. The application incorporates Jmol as visualization tool. Sanjeevini android application provides seemless switch between android devices by providing results on a centralized server. Results of a job can be accessed by multiple android devices by providing its unique JobID. This feature can be useful in sharing results and helping multiple researchers access and assess results relevent to collarating groups seperatedgeographically.The application, available free of cost, is expected to benifit researchers and acedemicians working in the field of drug discovery. It can be downloaded from SCFBio website (*http://www.scfbio-iitd.res.in*). Additionally, it will be made available shortly on google play store for wider audience coverge.

References

[1] Jayaram, B. *et al.* Sanjeevini: a freely accessible web-server for target directed lead molecule discovery. *BMC bioinformatics* 13, S7 (2012).

P-2

In silico based comparative study of archeal genome for mining of genes involved in phosphorus metabolism and stress tolerance

Alok Kumar Singh[1#*], Anil Kumar Singh[2], Alok Kumar Shrivastava[3]

[1,3] ICAR-National Bureau of Agriculturally Important Microorganisms, Kushmaur, Maunath Bhanjan, U.P., India – 275101; [2] Environmental Microbiology Laboratory, Environmental Toxicology Group; CSIR-Indian Institute of Toxicology Research (CSIR-IITR)

In present study, whole genome of two archea *Halolaminarubra* CBA1107 and *Halobacterium* sp. NRC-1 were retrieved from NCBI database. Functional based comparison between both genomes was carried using online server RAST (Rapid Annotations using Subsystems Technology). A total of 26 genes was present in both genomes that are responsible for phosphorus metabolism (18 genes) and stress tolerance (8 genes). Among genes involved in phosphorus metabolism, 5 genes (*Pst*S, *Pst*B, *Pst*A, *Pst*C, *Pho*U) were found in both genome but *Halolaminarubra*CBA1107 does have 4 genes encoding *Pst*S-Halobacteriales type, Alkaline phosphatase, Inorganic phosphatase, Alkaline phosphatase and low-affinity inorganic phosphate transporter. While, *Halobacterium* sp. NRC-1 possess 6 genes encoded glutaredoxin-related protein, glutamine-cysteine ligase archeal, glutathione S-transferase, lactoylglutathionelyase, superoxide dismutase and GTP-binding protein *Hfl*X that are

responsible for stress tolerance but *Halolaminarubra* CBA1107 does have any genes for stress tolerance.

P-3

Onco-Regulon: An Integrated database and software suite for site specific targeting of transcription factors

Akhilesh Mishra[1, 2], Shruti Aggarwal[1], Nirotpal Mrinal[3], B. Jayaram[1, 2*]

[1]Supercomputing Facility for Bioinformatics & Computational Biology, IIT Delhi,[2] Kusuma School of Biological Sciences, IIT Delhi, India, [3] Lab. of Molecular Biology, South Asian University, New Delhi, India.
**Corresponding author: B. Jayaram*
E-mail: bjayaram@chemistry.iitd.ac.in

Transcription factors (TFs) bind at multiple sites in the genome and regulate the expression of many genes. Regulating TF binding in a gene specific manner remains a formidable challenge in drug discovery because the same binding motif may be present at multiple locations in the genome. Here, we present Onco-Regulon (*http://www.scfbio-iitd.res.in/software/onco/NavSite/index.htm*), an integrated database of regulatory motifs of cancer geneswhich is clubbed with USP (**U**nique **S**equence-**P**redictor) a software suite that identifies unique sequences for each of these regulatory DNA motifs at the specified position in the genome. USP works by extending a given DNA motif, in 5′→3′, 3′→5′ or in both directions by adding one nucleotide at each step. This step is iterated till the frequency of the extended motif becomes unique in the genome. Thus, for each given motif we get three possible unique sequences of which the smallest motif can be further used as a drug target. The Closest Sequence Finder program will help in predictingthe off-target drug binding site in the genome. Inclusion ofDNA-Protein Structural Information will furthermakes Onco-Regulon a highly informative repository for gene specific drug development.

P- 4

Insight towards Protein Function Prediction using physicochemical parameters

#Amita Pathak[1, 2], Avinash Mishra[2, 3], Rahul Kaushik [2,3], B. Jayaram[1, 2, 3*]

[1]Department of Chemistry
[2]Supercomputing Facility for Bioinformatics & Computational Biology,
[3]Kusuma School of Biological Sciences, Indian Institute of Technology, HauzKhas, New Delhi-110016, India.
Email: bjayaram@chemistry.iitd.ac.in; amita@scfbio-iitd.res.in

Protein function prediction is a daunting task till date. The methodology reported here is based on machine learning approach where different physico-chemical properties of protein sequences were trained against their corresponding molecular functional classes. We have implemented the Random forest technique to build protein function prediction model. Four highly populated classes were selected from UniprotGO Annotations and given new names: Transporter, Molecular transducer, Receptors, Catalytic/Enzyme. If query protein sequence does not fall in any of the mentioned classes then output shows OTHERS; this means that protein can be from any other functional class except the mentioned ones. The features used for implementing the model are sequence length, hydrophobic residue, charged residues, polar residues, basic residues, aliphatic residue, molecular mass, sp^3 hybridized gamma carbon residues, short residues or absence of delta carbon, hydrogen bond donor residues, linear residues or absence of forks with hydrogen, cellular location, number of phosphorylation site, glycosylation site (N and C terminal), and occurrence of repeats. Average sensitivity and specificity observed are 0.82 and 0.75 respectively. Further improvements to the methodology are underway.

P- 5

ProTSAV+: A metaserver for identification and scoring of protein tertiary structures

Ankita Singh[1, 2#], Rahul kaushik[1, 3], Himani Kuntal[2], B. Jayaram[1, 3, 4*]

[1] Supercomputing Facility for Bioinformatics and Computational Biology, IIT Delhi, India
[2] Department of Bioinformatics, BanasthaliVidyapith, Banasthali, (India) - 304022
[3] Kusuma School of Biological Sciences, IIT Delhi, India
[4] Department of Chemistry, IIT Delhi, HauzKhas, New Delhi (India) – 110016

Statement of the Problem: Protein structure quality assessment is among the most important challenges in the field of structural biology. Recent methodological advancements in protein structure prediction approaches have created an immediate necessity for highly efficient quality assessment methods for discriminating good model structures. Better quality predicted protein structures may help in further biological function assignment and in structure based drug discovery.

Methodology & Theoretical Orientation: The ProTSAV+ metaserver integrates 11 individual approaches of quality assessment (Figure 1) and provides the user with a single quality score in case of individual model structure and ranking in case of multiple decoy structures. The ProTSAV+ performs weightage based combination of some of the widely used and thoroughly validated freely/ on request available tools. These tools mainly embed various structural and energetic features individually or in combination like accessible surface area, non-covalent interactions, residues based contact potentials etc..

Findings: The specificity and sensitivity of the metaserver are 88% and 91% respectively for good quality model structures. The same goes to 100% and 98% respectively for experimental structures. The updated version of the metaserver is fielded in the recently concluded CASP12 protein structure prediction experiment under 'QA category' wherein it performed well among some of the leading QA participants. For instance, the metaserver is ranked among top 20 participants in 22 targets out of 40 publically released targets.

Conclusion & Significance: The server overcomes the limitations of any single method and is seen to be robust in helping in improved quality assessment.

References

[1] Singh A, Kaushik R, Mishra A, Shanker A, Jayaram B (2016) ProTSAV: A Protein Tertiary Structure Analysis and Validation Server. BBA-Proteins and Proteomics 1864:11-19.

[2] Singh A, Mishra A, Khosravi A, Khandelwal G, Jayaram B (2016) Physico-chemical fingerprinting of RNA Genes", Nucleic Acids Research.doi:10.1093/ nar/gkw1236

[3] Jayaram B, Dhingra P, Mishra A, Kaushik R, Mukherjee G, Singh A and Shekhar S (2014) Bhageerath-H: A homology ab initio hybrid server for predicting tertiary structures of monomeric soluble proteins", BMC Bioinformatics 15(Suppl 16):S7

[4] Gupta DD, Kaushik R, Jayaram B (2016) Protein folding is a convergent problem!", Biochemical and Biophysical Research Communications, , 480 (4), 741-44.

[5] Mishra A *et al* (2014)"D2N: Distance to the native", BBA - Proteins and Proteomics, 1844 (10), 1798-1807.

P-6

Physico-Chemical Fingerprinting of RNA Genes

Ankita Singh[1#], Akhilesh Mishra[1, 2], Ali Khosravi[3], Garima Khandelwal[1,4] B. Jayaram[1, 2, 5*]

[1]Supercomputing Facility for Bioinformatics & Computational Biology, [2] Kusuma School of Biological Sciences, Ale-Taha Institute of Higher Education, Tehran, Iran, [3]Taha Institute of Higher Education, Tehran, Iran, [4] Cancer Research UK Manchester, The University of Manchester, Wilmslow Road, Manchester M20 4BX and [5]Department of Chemistry, Indian Institute of Technology, HauzKhas, New Delhi-110016, India

Over the years, various computational methods have displayed a potential for fast and accurate characterization of genes. Majority of these methods are knowledge-based and involve sophisticated statistical and mathematical techniques for training and prediction. An alternative approach to solve this complex challenge is based on the hypothesis that different functional units on genomic DNA differ in their physico-chemical properties, which, in principle, can be extracted from atomic models

of DNA. The present work encompasses elucidation of physico-chemical fingerprints for different functional units in prokaryotic and eukaryotic genomes on the basis of atomic level descriptions of oligonucleotides derived from molecular simulations. We advance here a novel concept for characterizing different classes of RNA genes on the basis of physico-chemical properties of DNA sequences. As knowledge-based approaches could yield unsatisfactory outcomes due to limitations of training on available experimental datasets, alternative approaches which utilize properties intrinsic to DNA are needed to supplement training based methods and to eventually provide molecular insights into genome organization.Based on a comprehensive series of molecular dynamics simulations of Ascona B-DNA consortium, we extracted hydrogen bonding, stacking and solvation energies of all combinations of DNA sequences at the dinucleotide level and calculated these properties for different types of RNA genes. Considering ~7.3 million mRNA, 255524 tRNA, 40649 rRNA (different subunits) and 5250 miRNA, 3747 snRNA, gene sequences from 9282 complete genome chromosomes of all prokaryotes and eukaryotes available at NCBI, we observed that physico-chemical properties of different functional units on genomic DNA differ in their signatures.

References

[1] Singhal,P., Jayaram,B., Dixit,S.B. and Beveridge,D.L. (2008) Prokaryotic gene finding based on physicochemical characteristics of codons calculated from molecular dynamics simulations. Biophys. J., 94, 4173-4183.

[2] Zhu,W., Lomsadze,A. and Borodovsky,M. (2010) Ab initio gene identification in metagenomic sequences. Nucleic Acids Res., 38, e132.

[3] Libbrecht, M. W. and Stafford W. (2015) Noble Machine learning applications in genetics and genomics Nature. Rev.Genet., 16,321-332.

[4] Stephs, Z. D., Lee, S. Y., Faghri, F., Campbell, R. H., Zhai, C., Efron, M. J. and Robinson, G. E. (2015). Big Data:Astronomical or Genomical? PLoS Biology., 13, e1002195.

P- 7

Insilco Study of Interaction of ZnO Nano cluster-Glucose oxidase; a Nano-bio-conjugate

Awadhesh Kumar Verma[1], Abhilash Jayaraj[2], Shashank Deep[2]*, Z. A. Ansari[1] and S.A. Ansari[1]

[1] Centre for Interdisciplinary Research in Basic Sciences, JamiaMilliaIslamia, New Delhi;
[2] Department of Chemistry, Indian Institute of Technology, New Delhi
**E-mail: sdeep@chemistry.iitd.ac.in*

To understand the mechanism of interaction between an enzyme and its substrate or with other molecule like its activators, inhibitors, or drugs have been crucial for biochemical research, but recently especially the interaction of nanoparticle with biomolecule has drawn enormous attentions of researcher in the field of nano-biotechnology, as it has future prospect of improving the entire healthcare processes for patient starting from diagnosis up to treatment, since it felicitates rapid testing and early diagnosis of the diseases. In present work, we have theoretically modeled 3-Dimensional cage like nano cluster of $(ZnO)_{12}$ in ground state configuration using Marvin Sketch software. We found that TRP 155 has maximum solvent accessible surface area i.e. 19.59 ($Å^2$), so $(ZnO)_{12}$ quantum cluster was docked to near TRP-155 of glucose-oxidase-FAD complex using autodockvina. Than we performed flexible docking of glucose with glucose oxidase using autodockvina, which shows that glucose was docked to near FAD molecule in the active site reported in literature. Glucose is making hydrogen bond with GOx. Further we performed docking of $(ZnO)_{12}$nano cluster with glucose oxidase, which shows that Lys-138, Trp-155, Cys-228, Gly-227, Cys-186, His-187, Asp-156, Ser-185 and Ala-160 of glucose oxidase involved in noncovalent interaction with $(ZnO)_{12}$. Finally, we performed the md simulation of $(ZnO)_{12}$nano cluster with glucose oxidase. Nano cluster is showing non-covalent interaction with Lys-138, Trp-155, Asp-156, Ala-159, Ser-160, Leu-163, Ser-185, Cys-186 and Cys-228 amino acid residues of GOx. The detail analysis of the interaction will be discussed at the time of presentation based on docking and MD-Simulation study.

P-8

Bhageerath H⁺: A software suite for tertiary structure prediction of globular proteins

Rahul Kaushik[1,2#], Ankita Singh[1], Debarati Das Gupta,[1,3] Shashank Shekhar[1] and B. Jayaram[1,2,3*]

[1]Supercomputing Facility for Bioinformatics & Computational Biology,[2] Kusuma School of Biological Sciences,[3]Department of Chemistry, Indian Institute of Technology, Hauz Khas, New Delhi (India)
[*]E-mail: bjayaram@chemistry.iitd.ac.in

The constantly widening gap between known protein sequences and experimentally solved structures, the need for structures of novel protein drug targets to benefit the structure based drug discovery have raised the necessity for developing highly reliable protein structure prediction approaches. Over the past two decades, a chronological assessment of improvements in structure prediction methodologies have been accounted via CASP experiments. The success of prediction methodologies has assisted the scientific community to explore the experimentally unsolved proteins more efficiently at their structural level. In the recently concluded CASP12 experiment, we tested BhageerathH+ software suite which delivers a reliable structure for the protein from sequence information. BhageerathH+ primarily comprises three major steps namely structure generation for conformational sampling, structure scoring for selecting the best conformations and structure refinement and side chain optimization for quality improvement. The sampled conformations (via Bhageerath, StrGen and NCL Align) are clustered for filtering out similar topologies. Post-clustering, the conformations are scored and ranked with an improved version of ProTSAV (ProTSAV+) for selecting top 50 conformations. The selected conformations are further processed with molecular dynamics based refinement and side chain optimization for enhancing their structural quality and re-ranked with ProTSAV+ for selecting top 5 conformations. The methodology was automated and fielded in the recently concluded CASP12 experiment under TS category as BhageerathH+ server wherein it is ranked jointly at 1st position for low resolution structures (under 7Å), 4th for medium resolution structures (under 5Å) and 8th for low resolution structures (under 3Å). BhageerathH+ software suite can be freely accessed at *http://scfbio-iitd.res.in/BhageerathH+*.

References

[1] Jayaram,B. et al. (2014). *Bhageerath-H*: A homology/ ab initio hybrid server for predicting tertiary structures of monomeric soluble proteins. BMC Bioinformatics**15(16),** S7.

[2] Jayaram, B. et al. (2006). *Bhageerath*: an energy based web enabled computer software suite for limiting the search space of tertiary structures of small globular proteins. *NAR34(21), 6195-6204.*

[3] Singh A. et al. (2015). ProTSAV: A Protein Tertiary Structure Analysis and Validation Server. *BBA - Proteins and Proteomics* 1864(1), 11-19.

P-9

Gut Microbiota: A future biomarker for obesity related diseases?

Shriya Madan[1 #], Suchita Reddy[1#], Anuvrat Sircar[1], Aakash T. Shaji[1], Avneesh Mittal[1], Purnima Anand[1], Neha Bansal[1], Balaram Pani[1], Ranjeet S. Thakur[1], Daman Saluja[3], Deepika Bhaskar[2] and Uma Chaudhry[1*]

[1] Bhaskaracharya College of Applied Sciences, University of Delhi, Delhi; [2] Deputy Dean Research, Vice-Chancellor's Office, University of Delhi, Delhi, [3] Dr. B. R. Ambedkar Centre for Biomedical Research, University of Delhi, Delhi
[*]E-mail: uma.chaudhry@bcas.du.ac.in

Obesity can simply be described as being the result of a long-term imbalance between energy intake and energy expenditure. It is turning out to be a serious public health issue affecting both children and adults. Prevention and management of obesity is proposed to begin in childhood when environmental factors exert a long-term effect on the risk for obesity in adulthood. Thus, identifying modifiable factors may help to reduce this risk. Although lifestyle is a major parameter influencing the condition, lot of studies are pointing out the link between obesity and gut microbiota composition. Further development in this area associates gut microbiome composition and maternal factors during pregnancy like probiotics, prebiotics, etc.

Method: A questionnaire about mother, child's prenatal and neonatal period, and early childhood was made. Healthy normal and obese individuals

participated and filled up the questionnaire. Stool samples were collected and used for 16S rRNA based detection and identification of predominant gut bacteria. Sequencing was carried out using Genotypic Technologies, Bangaluru facility.

Result: The metagenomic analysis of gut microbiome of obese and normal individuals identified organisms on the basis of V3 and V6 regions (variable regions of 16S rRNA). Further analysis of the data helped us in sorting out gut bacteria present in more prevalent manner in obese individuals. *Parabacteriodes* being one of them, was present in obese samples upto 46.9% whereas in normal samples they were just 3%. On similar lines, *Butyricimonas*was more prevalent in normal samples (about 18.9%) than obese (about 0.2%). Some of the bacteria were found to be exclusively present in obese samples, even though the their prevalence was low.

Conclusion: Impact of various maternal factors influences the gut microbiota. The more prevalent gut bacteria in obese individuals must contribute to obesity by upregulating some common pathways or adding some unique pathways. We may go further to analyse gut microbiota in patients having obesity related diseases. Possibility is that we may come up with some of these bacteria as biomarkers associated exclusively for obesity linked diseases. Outcome of the study would be non-invasive biomarkers for obesity associated diseases which could be evaluated early in life, during the childhood only and measures of control may be taken care of.

P- 10

Protein Folding Energetics analyzed through microsecond long Molecular Dynamics Simulations

#Debarati Das Gupta[1, 2], M. Varun[1], B. Jayaram[1,2, 3*]

[1]Department of Chemistry & [2]Supercomputing Facility for Bioinformatics & Computational Biology, [3]Kusuma School of Biological Sciences, Indian Institute of Technology, Hauz Khas, New Delhi-110016, India
[*]E-mail: bjayaram@chemistry.iitd.ac.in;
[#]E-mail: debarati@scfbio-iitd.res.in

Protein folding is considered as a grand challenge problem. The factors contributing to protein stability is still a mystery despite five decades of intensive research. A polypeptide chain is marginally stable by 5-15 kcal/mol and the resultant free energy can be parsed into individual components. In this work, we have collected the experimental free energy of folding for 20 proteins from the ProTherm Database. Using microsecond long MD simulations equivalent to five years of CPU time, we generated the native (N) ensemble and unfolded (U) ensembles and computationally calculated the free energy of folding for these 20 proteins. We then devised a protocol to decompose the net free energy in to enthalpic and entropic contributions. What we found is excellent correlation with experimental data (R^2= 0.75). We then utilized the nomenclature and knowledge of synthons and pattern recognition from small molecule crystallography; combined it with the structural and thermodynamic aspects of protein interactions and applied it to our trajectories of 20 systems. We queried whether there existed any universally transferable patterns of atoms common to all proteins, which accounted for all the crystallographically observed contacts. We find that there exists three 3-atom patterns which accounts for the backbone H bonding pattern as well as van der Waals packing in proteins which are termed as "foldable" synthons, essential for any protein to fold. This analysis generates a novel and simple view of intramolecular recognition and could at as the basic template for protein folding.

P-11

OTUX – A bacterial OTU database for accurate OTU-picking and easy cross-comparison of OTU profiles derived from different metagenomic studies

Deepak Yadav[#], Anirban Dutta, Sharmila Mande[*]

Biosciences R&D Division,TCS Research, Tata Consultancy Services Ltd., 54-B Hadapsar, Industrial Estate, Pune, Maharashtra, India, 411013

OTU-picking has become a de-facto standard for taxonomic analysis in 16S rRNAamplicon sequencing based metagenomics studies. OTU-picking can either be reference-based or *de novo*, both of which suffer from a few drawbacks. Given

that current methods for reference-based OTU identification or taxonomic classification rely on databases cataloguing full-length 16S rRNA genes or reference OTUs identified through clustering full-length 16S rRNA genes, results obtained using 'short-read' sequence queries (as generated by NGS platforms) during OTU identification/ classification can be inaccurate and sub-optimal. Further, rate of evolution (accumulation of mutations) is not always uniform across the length of a 16S rRNA gene across different taxonomic clades. It is possible that a short region remains identical during the course of evolution, whereas flanking regions are more prone to mutations or vice versa. Given these possibilities, *de novo* OTU clustering results (as well as downstream taxonomic classification) can significantly vary based on the specific region of 16S rRNA gene targeted for sequencing. Evidently, OTUs identified using reference-based vs de novo methods are also expected to provide different results given the above reasons.

To address these drawbacks we have created a 'customized' OTU database, OTUX, for different (hypervariable) regions of 16S rRNA gene. A taxonomic classification pipeline coupled to the OTUX database allows "open reference" OTU-picking using OTUX as reference. In addition, the results of OTU-picking, in terms of OTUX IDs, can be mapped back to popularly used OTU databases (such as Greengenes). This, in essence, allows for easy cross-comparisons between results of different metagenomics studies, that might have used different hypervariable regions and/ or different sequencing technologies. Validation studies performed with our method show that performing OTU-picking using OTUX and subsequently mapping back the results to conventional OTU databases, results in more efficient and accurate taxonomic classifications, compared to the conventional methods.

P-12

In silico identification and characterization of defense related Whirly transcription factor (WHY) in *Solanum melongena* L.)

Mohd Aamir[1], Dhiraj Mishra[2], Mukesh Meena[1], V. K. Singh[3], Shailesh Kumar Tiwari[4], B. Singh[4], Surendra Singh[1]*

[1]Department of Botany, Banaras Hindu University, Institute of Science, Banaras Hindu University, Varanasi- 221005; [2]Bachelor of Technology in Biotechnology, National Institute of Technology, Warangal-506004; [3]Centre for Bioinformatics, School of Biotechnology, Institute of Science, Banaras Hindu University, Varanasi- 221005; [4]ICAR- Indian Institute of Vegetable Research, Post Bag No. 01; P. O. Jakhani (Shahanshapur),Varanasi - 221 305

Whirly transcription factors (WHY) are the key players in regulating the plant defense mechanism. In this work, we have identified the putative WHY in *Solanum melongena* L. using the transcriptome sequences available at NCBI. The protein sequences available for whirly1 (4KOO), whirly2 (4KOP) in *Arabidopsis* and whirly2 (3R9Y) in *Solanum tuberosum* were retrieved for finding their sequential homologues in brinjal. The tBLASTn was done and the relevant similar sequences were retrieved from *S. melongena* transcriptome sequence assembly (TSA) database. The tBLASTn results identified 14 mRNA sequences, (accession ID GAYR01016883.1, GAYR01016885.1, GAYR01010421.1, GAYR01011122.1, GBEF01024573.1, GBEF01041696.1, GBGZ01050959.1, GBGZ01050960.1, GBGZ01050958.1, GBGZ01046356.1, GBGZ01050962.1, GBGZ01046355.1, GBGZ01050961.1 and GBGZ01046354.1) in *S. melongena* TSA database. In this work, we have focused on Ramnagar Giant, a local landrace of *S. melongena* with accession ID: GAYR01016883.1, GAYR01016885.1, GAYR01010421.1 and GAYR01011122.1. The results indicate availability of only two types of WHY in Ramnagar Giant. SmWHY1 (GAYR01010421.1) and SmWHY2 (GAYR01016883.1) genes were successfully identified in Ramnagar Giant. Further, phylogenic analysis of both SmWHY proteins was done to find out the phylogenetic relationships of Ramnagar Giant with seven other Solanaceae family members (*Solanum tuberosum, Solanum lycopersicum, Solanum penellii, Nicotiana tomentosiformis, Nicotiana attenuate, Nicotiana tabacum and Capsicum*

annuum).The phylogenetic classification revealed that SmWHY1 was closely similar to *S. tuberosum, S. lycopersicum, S. penellii* and *C. annuum*. In the case of SmWHY2, *S. tuberosum, S. lycopersicum* and *S. penellii* were found to be closely similar. In diversity analysis, whereas SmWHY1 was found to be diversified from *N. tomentosiformis, N. attenuate* and *N. tabacum*, in case of SmWHY2, *N. tomentosiformis, N. attenuate, N. tabacum* and *C. annuum* were, however, showing diversity. The swiss modeller was used for structural modelling. Further the protein models generated were further evaluated for their qualitative and quantitative analysis, model structure reliability and stability. The predicted models were found to be good enough on quality checkup and satisfied all the crucial energy parameters. The functional annotation in terms of gene ontological enrichment analysis was carried out using CATH REViGO web server. The REViGO analysis revealed that the protein have been associated with DNA binding, RNA binding and single stranded DNA binding.

P-13

Computational Analysis of Autoimmune Lymphoproliferative Syndrome (ALPS) Causal Genetic Variants of Fas and FasL – A structural insight into Death Receptor/ Ligand deformities

Saida Sadath[1, 2], Kaleemuddin Mohammed[1,2#*], Furkhan Mohammed[3], Babajan Banaganapalli[1,4], Nihal Hasan Mohammed[2], Noor Ahmad Shaik[1,4], Fahad A. Al-Abbasi[2]

[1]Princess Al-Jawhara Al-Brahim Center of Excellence in Research of Hereditary Disorders (PACER-HD), King Abdulaziz University, Jeddah, KSA
[2]Department of Biochemistry, Faculty of Science, King Abdulaziz University, Jeddah, KSA
[3]Department of Biology, Faculty of Science, King Abdulaziz University, Jeddah, KSA
[4]Department of Genetic Medicine, Faculty of Medicine, King Abdulaziz University, Jeddah, KSA
kaleem_kamran111@yahoo.com

Genetic variations in Fas receptor and its ligand are observed in wide spectrum of immune system disorders, most prominently causing Autoimmune Lymphoproliferative Syndrome (ALPS). Till date, more than 300 families with approximately 500 patients were diagnosed with hereditary ALPS, worldwide over the last 20 years [1]. Nevertheless, the underlying basis of how these deleterious genetic variations elicit protein conformation in terms of structural stability, and protein binding affinity remains unexplored. Therefore, we aimed to study the structural and functional impacts of Fas/FasL mutations, by *in-silico* method as an alternate to traditional *in-vivo* and *in-vitro* approaches. Fas/FasL genetic variations details were collected from different databases and their corresponding clinical associations were confirmed by the text-mining. Initially, various computational algorithms were employed to categorize the genetic variations based on the degree of deleterious nature to Fas/FasL protein structures. Then the *ab-initio* protein structures for Fas/FasLwildtypes and mutant models of the most deleterious mutations were built by I-Tasser server and Swiss Model-Expasy respectively. Molecular docking was also performed to assess the binding affinity of wildtype and mutated protein of Fas/FasL complex. We were able to identify five genetic variations mapped on highly conserved death domain region (exon-9) of Fas inducing significant conformational changes in the mutant proteins which alter the stability of Fas-FasL interactions resulting in ALPS. This study supports *in-silico* approach as a primary filter to verify the plausible degree of deleterious mutations based on the evolutionary conservation of sequence, structural homology and protein stability.

References

[1] Li P, Huang P, Yang Y, Hao M, Peng H, Li F. Updated understanding of autoimmune lymphoproliferative syndrome (ALPS). *Clin. Rev. Aller. &Immuno.* 2016 Feb 1;50(1): 55-63.

P- 14

Correlating Substituent Structure-Solubility Relationship to Design Prodrug with Improved Solubility Profile

Nupur S Munjal[a#], Manu Sharma[b], Chittaranjan Rout[a*]

[a]Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology, Waknaghat, Solan, Himachal Pradesh, 173234, India; [b]College of Pharmacy, Maharishi Markendeshwar University, Mullana-Ambala, Haryana, India
E-mail: chittaranjan.rout@juit.ac.in

Paclitaxel (taxol) is a significant anti-cancer drug, but its use is limited mostly due to poor aqueous solubility and bioavailability [1]. Prodrugs, compounds that undergo bio-transformations before showing pharmacological effects, were developed to increase the solubility; however, no rational approach is followed. As substituent groups contribute significantly to solubility, the current study implemented quantitative structure property relationship (QSPR) model to correlate structural property of Paclitaxel substituents with solubility of prodrugs. Mainly 2′- and 7′-sites substitutions were reported for Paclitaxel prodrugs [2]. As Paclitaxel group is common in all prodrugs, the current study aims to develop QSPR model which correlate structure of substituent (2′-site) with solubility. Geometry optimization of substituent was performed at PM6 level using Gaussian software. Thirty groups of descriptors were extracted for each substituent using Dragon7 software. The selection of descriptors was performed in MATLAB using 'stepwise fit'. Only few descriptors from some groups provided good regression (R^2) and correlation (Q^2) coefficients. After evaluating many combinations, fifteen descriptors from 2D-Autocorrelation, Burgeneigen-values, 2D Atoms-pair groups provided R^2 and Q^2 values 0.89 and 0.88, respectively. This model performance was also evaluated on substituent groups present at 7′-site of Paclitaxel prodrugs. The R^2 value was 0.79 which indicates that this model can be used for solubility prediction. This approach can assist synthetic chemistry to make structural modifications that may improve solubility profiles of prodrugs.

Refrences

[1] U. Yasutsugu, B. M. Amarendta, J. O. Knipe, William C. Rose, A. M. Casazza and M. V. Dolatrai , " Novel water soluble phosphate prodrugs of taxol possessing in vivo antitumor activity", Bioorganic & Medicinal Chemistry Letters, Vol.3, No.8, pp. 1761-1766, 1993.

[2] M.Ziyuan, L. Quanxia, L.Jun, Y.Houzong, L. Xiaoqing, J. Feng Jiang, L. Aiping and Z. Ge, "Prodrug strategies for paclitaxel", International Journal of Molecular Sciences, Vol. 17, No. 5, pp 796, 2016.

P-15

An in silico approach to discover lead molecules targeting DNA unique to pathogens

Akhilesh Mishra[1, 2, #], Pradeep Pant[1, 3, #], Nirotpal Mrinal[4] and B. Jayaram[1, 2, 3, *]

[1]Supercomputing Facility for Bioinformatics & Computational Biology, Indian Institute of Technology Delhi, [2]Kusuma School of Biological Sciences, Indian Institute of Technology Delhi, [3]Department of Chemistry, Indian Institute of Technology Delhi, and [4]Laboratory of Molecular Biology, South Asian University, New Delhi.
[*]E-mail: bjayaram@chemistry.iitd.ac.in
[#] These authors have contributed equally

With the rapid emergence of antimicrobial resistance, development of alternative medications for new drug targets to overcome this crisis is the prime focus of several academic laboratories and pharmaceutical companies. Here we report a step by step computational protocol for identifying DNA sequences unique to pathogens and absent in the host and non-pathogenic strains of the microbe, and for targeting these sequences with small molecules via virtual screening against a million compound library to identify good candidate molecules which can bind to these unique DNA targets with high affinity. This methodology is demonstrated on Mycobacterium tuberculosis H37Rv, wherein a new octamer unique to pathogenic strain has been identified and a few hit molecules have been proposed. Being fast and cost effective, this protocol could be of importance in generating new potential drug candidates against infectious organisms for further experimental studies. This methodology is freely available at *http://www.scfbio-iitd.res.in/PSDDF/.*

References

[1] Mishra, A.; Pant, P.; Mrinal, N.; Jayaram, B. A Computational protocol for the discovery of lead molecules targeting DNA unique to pathogens, *Methods*, 2017, *https://doi.org/10.1016/j.ymeth.2017.07.017*

P-16

Genome to Hit molecules case study of Dengue Virus using Dhanvantari Pipeline

Ruchika Bhat[1,2] and B Jayaram[1,2,3]*

[1]Department of Chemistry; [2]Supercomputing Facility for Bioinformatics & Computational Biology; [3]Kusuma School of Biological Sciences; Indian Institute of Technology, Hauz Khas, New Delhi-110016, India
Email: bjayaram@chemistry.iitd.ac.in; ruchika@scfbio-iitd.res.in

The Broad Viral Genomics Group has focus on four highly classified viruses namely HCV, HIV, Dengue, and WNV. However, in case of Dengue virus a 30-fold upsurge worldwide between 1960 and 2010 is reported and about 500000 persons are hospitalized every year with 2.5% mortality rate. No drug against Dengue infection has yet been approved. In an attempt to combat the disease, the automated pipeline Dhanvantari was explored. The repurposing of the already known FDA approved drugs was performed to speed up the drug discovery against the infection. The mechanism of action was also studied through docking and scoring studies. There were 58 ORFs identified for the genome sequence of Dengue Virus 2 (NC_001474.2) from the pipeline. The best five ORFs were considered for 3D structure prediction. The active sites of the same were identified using the Drug module of Dhanvantari pipeline and the best screening hits were further docked to refine results for more precision. The study showed Etravirine, Hexafluronium, Ambenonium and Spiramycin as best hits against the RNA dependent polymerase of virus which showed no similarity with human proteome and hence, considered as a potential drug target. These FDA drugs can further be investigated with *in vitro* studies to correlate their inhibitory effect.

P-17

Hepatitis A cysteine protease antagonists: Design, synthesis and study of the mechanism of inhibition

Ruchika Bhat[1,2], Sohona Gangopadhyay[4], Kartik Lakshmi Rallapalli[1], Kamalika Banerjee[3], V. U. Bhaskara Rao[1], R P Singh[1], Manidipa Banerjee[3], B Jayaram[1,2,3]*

[1]Department of Chemistry, Indian Institute of Technology, Delhi, Hauz Khas, New Delhi-110016, India, [2]Supercomputing Facility for Bioinformatics & Computational Biology, Indian Institute of Technology, Delhi, Hauz Khas, New Delhi-110016, India
[3] Kusuma School of Biological Sciences, Indian Institute of Technology, Delhi, Hauz Khas, New Delhi-110016, India
[4] Geological Survey of India, Western Region, Jaipur, India

World has seen Hepatitis A as a jaundice causing virus, but this virus can become a potential threat to the mankind with its less intricate yet non treatable form. Yearly 1.4 million cases of the HAV infection are reported with a mortality of up to 2.5%. In order to contemplate the complete eradication of this disease, a potent drug to treat the people infected with HAV needs to be developed. The HAV life cycle is dependent on its cysteine protease; 3Cpro, which cleaves the polyprotein into functional proteins. The vitality of protease enzyme in the viral lifecycle makes it an attractive drug target. In this study we have designed a number of small molecule inhibitors against the 3Cpro via computational methods. A plausible mechanism of action has been elucidated via exhaustive computational approaches. MMGBSA analysis along with RMSF calculations and C-alpha distance mapping suggest that these inhibitors laterally move towards the active site (HIS44 and CYS172) thereby blocking its functionality, thus specifically inhibiting the enzyme. These inhibitors show promising antiviral properties with micro molar range inhibitions in *in vitro* assays against 3Cpro enzyme with least toxicity values.

P-18

Symmetric nucleic acids

#Pradeep Pant[1, 2] and B. Jayaram[1, 2, 3*]

[1]Department of Chemistry
[2]Supercomputing Facility for Bioinformatics & Computational Biology,
[3]Kusuma School of Biological Sciences,
Indian Institute of Technology, Hauz Khas
New Delhi-110016, India
[*]E-mail: bjayaram@chemistry.iitd.ac.in
[#]pradeep@scfbio-iitd.res.in

Directionality (5′→3′) is fundamental to the nucleic acid architecture and is essential for replication and transcription. We observed that this directionality can be manipulated either by breaking (C3′ to C2′) or making (C5′ to C2′) a chemical bond in each nucleotide unit leading to symmetric nucleic acids. We computationally designed a few novel symmetric nucleic acids and investigated their conformational stability and flexibility via detailed all atom explicit solvent 100-ns long molecular dynamics simulations and observed that some of them retain the regular double helical B-DNA structure. Besides stimulating academic curiosity, we believe that these symmetric nucleic acids could be of potential synthetic and therapeutic value on par with the peptide nucleic acids

References

[1] P. Pant, S. A. Shaikh, B. Jayaram, Biopolymers 2017, 107(4).

[2] J. D. Watson, F. H. C. Crick, Nature 1953, 171, 737.

[3] G. Ravishanker, P. Auffinger, D. R. Langley, B. Jayaram, M. A. Young, D. L. Beveridge, Rev. Comput. Chem. 1997, 11, 317.

[4] S. A. Shaikh, B. Jayaram, J. Med. Chem. 2007, 50, 2240.

P-19

A transfer-of-training framework for similarity detection and classification of protein-protein complexes

Shruti Gupta[1#], Manisha Kalsan[1], Dana Mary Varghese[1], Ajay Arya[1], Ajay Kumar Verma[1], Shandar Ahmad[1*]

[1]School of Computational and Integrative Sciences, Jawaharlal Nehru University, New Delhi, India
** Corresponding author E-mail: shandar@jnu.ac.in*

Protein-protein interactions (PPIs) form the basic building blocks of all biological functions in higher organisms. High throughput techniques such as mass spectrometry have unraveled a large number of these interactions. However, the enormity and complexity of PPI throws up a challenge to the understanding of their roles in a given biological context. Classification of large populations in a data set is a way to understand collective behavior of entities in a simpler and hierarchical manner. Many sequence and structure based methods of classifying single proteins are available but the work on classifying complexes themselves is far more challenging and lags behind. Sequence and structural basis of such a classification are not fully developed and it is not clear if they would capture the true functional landscape of PPIs. In this work, we present a novel framework to estimate the similarity between protein-protein complexes. These estimates are based on the concept of transfer of training (TOT), under the intuitive assumption that even the apparently unrelated complexes share common features, which are generalized by a computational model during training. Ability of a model to predict binding sites will be enhanced if a similar complex is present in the training data set from which this model was derived. Thus, the similarity between two complexes A and B can be estimated by taking the difference between the predictability of binding sites in A obtained from models trained with and without the data from the complex B. As a converse application, we also show that the TOT similarity between complexes is also weakly related to the summary of the evolutionary profiles of constituent proteins, which can thereby be used to improve the performance of binding-site prediction methods.

References

[1] Ahmad, S and Mizuguchi, K Partner-Aware Prediction of Interacting Residues in Protein-Protein Complexes from Sequence Data PloS one 6 (12), e29104 (2011).

[2] Jordan, R.A. and Yasser, E.L.M. and Dobbs, D. and Honavar, V. *et al.,* Predicting protein-protein interface residues using local surface structural similarity, BMC Bioinformatics, 13, 41, (2012).

[3] Schwikowski, Benno, Peter Uetz, and Stanley Fields. A network of protein-protein interactions in yeast. Nature biotechnology 18, 1257-1261 (2000).

P-20

Expression and purification of Fibulin-5 in *E. coli*: A highly protease speceptible protein

Sweta Singh#, Shradha Jamwal, Jaideep Kumar, Sudarshan Kumar*, A. K. Mohanty

Email: swetasinghh93@gmail.com, ishuangel17@gmail.com, jai.deeshu@gmail.com, ashokmohanty1@gmail.com

Fibulin-5 (FBLN5) also known as short fibulin, is isoform of fibulin family proteins. Recently, this molecule has been proven to be a potential biomarker in various types of cancers and early detection of pregnancy in cows and buffaloes. In this study production of recombinant Buffalo FBLN5 was attempted in *E. coli* and a novel method was developed to enhance the production of pure protease susceptible BuFBLN5. The 1275 bp BuFBLN5 gene (without signal peptide) was cloned pET22b(+) vector and Lemo21(DE3) bacterial cells were transformed with pET22b(+)_BuFBLN5. Expression conditions were optimized in *E. coli* to find optimum temperature, IPTG concentration, protease inhibitors for the production of full length 55 kDa BuFBLN5 in soluble form. Ni-NTA based purification resulted in production of cleaved products of 15 kDa, 23 kDa, 38 kDa, 55 kDa and 70 kDa which were confirmed by Anti-His tag antibody and mouse origin human Anti-Fibulin-5 monoclonal antibody. The obtained cleaved products were of C-terminal as it was detected by anti-fibulin-5 monoclonal antibody generated against C-terminal portion of the protein and the His tag associated at the C-terminal of the recombinant BuFBLN5.

P- 21

Design and validation of novel antagonists of stromelysin-1 (MMP-3) through core hopping, ROC metrics and molecular dynamics simulations

Sudheer Kumar Katari#, Chiranjeevi Pasala, RavinaMadhulithaNalamolu, Amineni Umamaheswari*

*Bioinformatics Centre, Department of Bioinformatics, SVIMS University, Tirupati-517507, *E-mail: svims.btisnet@nic.in*

Matrix metalloproteinase-3 (MMP-3) or stromelysin-1 is a zinc dependent endopeptidase involved in tumor metastasis and malignancies of breast, lung and pancreas etc. Hyper activated human stromelysin-1 was involved in tumor microenvironment niches and metastasis. MMP-3 proteinase activates other MMP pro-enzymes (GelatinaseA-MMP-2, Matrilysin-MMP-7 and GelatinaseB-MMP-9) which degrade ECM, mainly II, III, IV, IX, X collagens, proteoglycans, fibronectins, laminins and elastins. Thirty one crystal structures of MMP-3 are available in the protein data bank, among which twenty four have diverse co-crystal ligands are retrieved for structure based rational drug design. Twenty four diverse co-crystal ligands were docked by rigid receptor docking (RRD) and quantum polarized ligand docking (QPLD) and binding free energies were calculated by Prime-MM/GBSA approach (Madhulitha *et al*., 2017) with the best resolute structure (1HY7). Core hopping technique was applied to L04, the best docked co-crystal ligand to generate novel compounds from the available 4 core libraries using Phase Shape's efficient algorithm. Three compounds have better binding affinities in multiple docking strategies (RRD and QPLD) and binding free energy calculations compared to the twenty four co-crystal ligands. Further the three compounds were validated with parent compound and three compounds with the 24 co-crystal ligands by receiver operative characteristic curve metrics (ROC: 0.9995 and ROC: 0.849056) in the presence of thousand decoy molecules obtained from directory of useful decoys (DUD) dataset. Thus, three compounds were validated as leads. Lead1 and L04 have QPLD XP GScore of -15.169 kcal/mol and -13.588 kcal/mol; ΔGScore of -114.418 kcal/mol

and -72.734 kcal/mol respectively. Stability of MMP-3-lead1 and MMP-3-L04 complex in natural physiological conditions were analyzed by molecular dynamics simulations using Desmond v4.3 for 50 ns (Katari*et al.*, 2016). Lead 1 is forming ionic bonds with catalytic residues His701, Glu702, His705, His711 and Zn. Hence, lead1 ceases the catalysis of MMP-3 by inhibiting the protease activity which in turn inactivates other MMP zymogens involved in metastatic progression of various cancers.

References

[1] Ravina Madhulitha N, Pradeep N, Sandeep S, Hema K, Chiranjeevi P, Sudheer Kumar K and Umamaheswari A. E-Pharmacophore Model Assisted Discovery of Novel Antagonists of nNOS. Biochemistry and Analytical Biochemistry. 2017; 6(1): 1-9.

[2] Sudheer Kumar K, Pradeep N, Sandeep S, Hema K, Chiranjeevi P and Umamaheswari A. Inhibitor design against JNK1 through e-pharmacophore modeling docking and molecular dynamics simulations. Journal of Receptors and Signal Transduction. 2016; 36(6): 558-571.

P-22

In-silico analysis of receptor with ovalbumin in asthma

Taru Aggarwal[1#], Ridhima Wadhwa[1], and Kumud Bala[*]

[1]Amity Institute of Biotechnology, Amity University, Uttar Pradesh, India

Role of Ovalbumin as a possible allergen causing Asthma was assessed. The rats were exposed once daily to aerosol of ovalbumin (1%, in 0.8% Saline) for 10 minutes followed by rest for 2 hours and then exposed to adjuvant (alum) (1% solution, 1ml/min) for 10 minutes. This procedure is repeated for 10 days. Later on adjuvant aerosol is discontinued whereas ovalbumin exposure is continued till 21 days. On 22nd day the rats was examined and samples for sensitized animals were collected. 10 rats were are further challenged using ovalbumin aerosols for 3days per week for 5weeks. Serum and BAL fluid was examined for lipid peroxide (MDA), reduced glutathione (GSH) and nitric oxide (NO) scavenging activities.

In the above data it has been observed that there is a significant change in level of MDA in both sensitized and challenged rats. The data for sensitized and OVA challenged rats in both serum and BALF samples were compared. Blood serum levels of rats showed significant decrease of GSH in OVA challenged rats as compared to sensitized rats. As per as nitric oxide free radical scavenging activity is concerned It has been found to increase in both Serum and BAL fluid.

In silico analysis using HEX, version 8.0.0, was done for ovalbumin as ligand and Asthma receptors obtained from Drug Bank. Docking score was found that few structures of protein kinase C receptor (alpha type, gamma type, theta type, zeta type) can bind ovalbumin. It was found that PDB IDs 2ELI, 2E73, 2ENZ and 2MJZA have minimum binding energies as compared to other molecules. Protein Kinase Theta type receptor was found to possess minimum binding energy.

P-23

Computer based design, synthesis and biological evaluation of some new quinoline based *Plasmodium falciparum* phophoethanolamine methyl transferase (*Pf*PMT) inhibitors

Anju Singh[1#,] Ashutosh Shandilya[2], Ashok Patel, [2] Shashank Deep[2] Nasimul Hoda,[1] and B. Jayaram[2*]

[1]Drug Design and Synthesis Lab, Department of Chemistry, Jamia Millia Islamia (Central University), New Delhi 110025, India
[2]Department of Chemistry, Indian Institute of Technology Delhi, Hauz Khas, New Delhi-110016, India
E-mail: nhoda@jmi.ac.in

Malaria is a major life threatening disease caused by *plasmodium parasites. Plasmodium falciparum* phosphoethanolamine methyltransferases (*Pf*PMT) plays a critical function in parasite development and differentiation. Here we designed and synthesized a series of quinoline based *Pf*PMT inhibitors. Molecular docking and scoring techniques were utilized for the design of inhibitors. A plausible mode of action for most active molecules was also proposed via molecular dynamics simulation approach. The molecules were

synthesized via an easily accessible, convergent synthetic route. All the synthesized molecules were assayed in the standard *P. falciparum* for their ability to inhibit the growth of cultured parasites and were found to have IC_{50} values in a nano molar range.

References

[1] Reynolds, J. M.; Takebe, S.; Choi, J.-Y.; El Bissati, K.; Witola, W. H.; Bobenchik, A. M.; Hoch, J. C.; Voelker, D. R.; Mamoun, C. B. J. Biol. Chem. 2008, 283 (12), 7894-7900.

[2] Shandilya, A.; Hoda, N.; Khan, S.; Jameel, E.; Kumar, J.; Jayaram, B. J. Mol. Graph. Model. 2017, 71, 96–103.

[3] Ajroud, K.; Sugimori, T.; Goldmann, W. H.; Fathallah, D. M.; Xiong, J.-P.; Arnaout, M. A. J. Biol. Chem. 2004, 279 (24), 25483–25488.

P-24

Interaction of Human Growth Hormone (HGH) and Penicillin Binding Protein (PBP), Serine Carbapenemase through molecular docking and computational tools

Mohit Kumar[*#] and Kushneet Kaur Sodhi, Pallee Shree, Dileep Kumar Singh

Soil Microbiology and Environmental Toxicology Laboratory Department of Zoology, University of Delhi, Delhi-110007, India
[*]E-mail: mohitzoologydu@gmail.com

Somatotropin a human growth hormone (HGH), which is secreted from somatotropic cells of anterior pituitary under the influence of growth hormone releasing hormone (GHRH) from the hypothalamus. In this study possibility of the growth hormone interaction with the antibiotic resistant bacterial enzymes in the children who are on medication and residing in slums and poor hygienic conditions were checked. Their interaction might be influence the growth of children during their growth period. This study showed that interaction between antibiotic resistant enzyme serine carbapenemase (a penicillin binding protein) and human growth hormone through molecular docking using HADDOCK server. 3D structure of both kind of proteins and their validation was also done through the Ramachandran plot. The study revealed, the specific amino acids for both the protein models, which might be involved in protein-protein interaction. Models were further validated by the proSA software for Z-score plotted on the basis of X-ray and NMR. By this study, we can conclude that, during the childhood, keep away the children from medicinal and all kinds of drugs to resist their growth inhibition.

P-25

Rational Design, Synthesis and Biological Screening of Triazine-triazolopyrimidine Hybrids as Multitarget Anti-Alzheimer Agents

Ehtesham Jameel[1#], Poonam Meena[2], Manisha Tiwari[2*] and Nasimul Hoda[1*]

[1]Department of chemistry, Jamia Millia Islamia, New Delhi-110025, India
[2]Dr. B. R. Ambedkar Centre for Biomedical Research, University Delhi, New Delhi 110007, India

In our endeavor towards the development of potent multitargeted ligands for the treatment of Alzheimer's disease, a series of triazine-triazolopyrimidine hybrids were designed, synthesized and characterized by various spectral techniques. Molecular docking tools were used to design the molecules and were synthesized via feasible convergent synthetic routes. In total, seventeen compounds were synthesized in which the di-substituted triazine-triazolopyrimidine derivatives (9a-b) showed better acetylcholinesterase (AChE) inhibitory activity than the corresponding tri-substituted triazine-triazolopyrimidine derivatives (10a-b). Interestingly, 9a and 9b demonstrated good inhibition selectivity towards AChE over BuChE by ~28 folds. In silico ADMET profiling highlighted these novel triazine derivatives have appropriate drug like properties and possess very low toxic effects in the primarily pharmacokinetic study. Overall, the multitarget profile exerted by these novel triazine molecules qualified them as potential anti-Alzheimer drug candidates in AD therapy.

P-26

Design, synthesis and biological screening of cyanopyridine-triazine hybrids as lead multitarget anti-Alzheimer agents

Mudasir Maqbool[1#], Manisha Tiwari[2], Nasimul Hoda[1*]

[1]Drug Design and Synthesis Lab, Department of Chemistry, Jamia Millia Islamia (Central University), New Delhi 110025, India
[2]Dr. B.R. Ambedkar Center for Biomedical Research, University of Delhi, Delhi
[*]E-mail: nhoda@jmi.ac.in

A series of new cyanopyridine-triazine hybrids were designed, synthesized and screened as multitarget anti-Alzheimer's agents. These molecules were designed while using computational techniques and were synthesized via a feasible concurrent synthetic route. Inhibition potencies of synthetic compounds 4a-4h against Cholinesterases, $A\beta_{1-42}$ disaggregation, oxidative stress, cytotoxicity, and neuroprotection against $A\beta_{1-42}$-induced toxicity of the synthesized compounds were evaluated. Compounds 4d and 4h showed promising inhibitory activity on acetylcholinesterase (AChE) with IC_{50} values ranging from 0.059 and 0.080 μM respectively, along with good inhibition selectivity against AChE over butyrylcholinesterase (BuChE). Molecular modelling studies revealed that these compounds interacted simultaneously with the catalytic active site (CAS) and the peripheral anionic site (PAS) of AChE. The mixed type inhibition of compound 4d further confirmed their dual binding nature in kinetic studies. Furthermore, the results from neuroprotection studies of most potent compounds 4d and 4h indicate that these derivatives can reduce neuronal death induced by H_2O_2-mediated oxidative stress and $A\beta_{1-42}$ induced cytotoxicity. In addition, *in silico* analysis of absorption, distribution, metabolism and excretion (ADME) profile of best compounds 4d and 4h revealed that they have drug like properties. Overall, these cyanopyridine-triazine hybrids can be considered as a candidate with potential impact for further pharmacological development in Alzheimer's therapy.

References

[1] M. Maqbool *et al., Eur. J. Med. Chem.* 2016, *107*, 63–81.
[2] M. Maqbool, *et al., Bioorg. Med. Chem.* 2016, 24, 2777–2788.

P-27

Identification of Putative Drug Targets in *C. albicans* through Subtractive Genomics, Metabolic Pathway and Gene Network Analysis

Rashi Verma[1,2#], Dibyabhaba Pradhan[2], Saumya Chaudhary[2,3], Arun Kumar Jain[2*], Luqman Ahmad Khan[1*]

[1]Department of Biosciences, Jamia Millia University, New Delhi – 110025, [2]Biomedical Informatics Centre, National Institute of Pathology, ICMR, New Delhi, India – 110029, [3]Dept. of Molecular & Cellular Engineering, Sam Higginbottom University of Agriculture, Technology &Sciences, Allahabad, India - 211007
*Email: #rv.01.nip@gmail.com, *drakjain@gmail.com, *lkhan@jmi.ac.in*

The polymorphic fungus Candida albicans is a member of the normal human microbiome and is used as a model organism for biology. In most individuals, *C. albicans* resides as a lifelong, harmless commensal. Under certain circumstances, however, *C. albicans* can cause infections that range from superficial infections of the skin to life-threatening systemic infections (Sardi et al.,2013). Comparative genomics has off late out-focussed the traditional methods of drug discovery. In this computational approach, whole proteome of *C. albicans* were retrieved from NCBI database, which was then analysed in CD-hit suite to eliminate the paralogous proteins followed by similarity search against Database of essential genes for identification of essential proteins. Proteins involved in unique pathways were analysed using Kyoto Encyclopedia of Genes and Genome (KEGG) database and Cytoscape. Whole proteome of *C. albicans* consists of 5000 proteins; out of which 2299 proteins were filtered as non-homologous to human.Essentiality analysis has showed 120 proteins can be considered as putative drug targets. These targets are identified to play vital role in survival and pathogenesis of *C. albicans*. Prioritization of these targets through choke point analysis, gut flora non-homology, druggabilityanalysis further established potential

of these proteins as drug targets. Ten hub proteins are proposed as target of interest based on high betweenness centrality, degree of nodes and closeness centrality. Therefore, these proteins could be ideal target of choice for developing potent future therapeutics against *C. albicans* infections with subsequent experimental validations.

References

[1] Sardi JC, Scorzoni L, Bernardi T, Fusco-Almeida AM, Giannini MM. Candida species: current epidemiology, pathogenicity, biofilm formation, natural antifungal products and new therapeutic options. Journal of medical microbiology. 2013; 62(1): 10-24.

P-28

An Information System on Buffalo (*Bubalus bubalis)* Genome

Amit Kairi[1#], Tanmaya Kumar Sahu[1*], Dr. Atmakuri Ramakrishna Rao[1*]

[1]Indian Agricultural Statistics Research Institute
Email: amit.kairi90@gmail.com, tanmayabioinfo@gmail.com, rao.cshl.work@gmail.com

Among the livestock species, buffalo has remained an integral part of the Indian rural economy. Significant amounts of information on different functional elements of various breeds of buffalo genome are available in public domain. However, the annotations of functional elements on the genome and the 3D structures of buffalo proteins are not available and there exist no browser available to visualize the genic information on buffalo genome. Hence, a study has been taken up with the objectives to (i) retrieve and the nucleotide and protein sequence information available in public domain and thereby develop a database (ii) develop user-friendly Web-based information system (iii) develop a genome browser to map gene information on genome. The nucleotide and protein sequence information has been retrieved from NCBI and parsed. Later on, the 3D structures of proteins were predicted, validated, refined and stabilized. A database was populated with the aforementioned information. An information system on buffalo genome with 3-tier architecture has been developed with MySQL database as bottom layer, Personal Home Page (PHP) as server side application-middle layer, HTML, CSS and JavaScript at top layer as client side application layer (URL: *http://cabgrid.res.in:8080/bgis*). The developed information system contains 930 complete CDS, 1154 partial CDS, 656 Exon sequences, 237 Intron sequences, 1709 Mitochondrial DNAs, 73 sequences of Promoter Region and 67 sequences of UTR region. Out of 930 buffalo genes, 837 have been found to be mapped onto *Bubalus bubalis* (female Murrah breed). The developed genome browser shows that maximum number of genes are distributed on chromosome 4 followed by chromosome 18. The results from the study also reveal that 837 out of 930 genes are mapped onto *Bubalus bubalis* genome. Whereas, 561 buffalo genes are mapped onto *Bos taurus* genome. Further, 202 genes are found to be predicted as orthologues between cattle and buffalo genomes.

References

[1] Iannuzzi, L (2007) The water buffalo: Evolutionary, clinical and molecular cytogenetics. *Italian Journal of Animal Sciences*: 2(supp2), 227-236.

[2] Sugawara H, Ikeo K, Fukuchi S, Gojobori T, Tateno Y. (2009) DDBJ dealing with mass data produced by the second generation sequencer. *Nucleic Acids Research*.**37**: D16-8.

[3] Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, Church DM, DiCuccio M, Edgar R, Federhen S, Geer LY, Kapustin Y, Khovayko O, Landsman D, Lipman DJ, Madden TL, Maglott DR, Ostell J, Miller V, Pruitt KD, Schuler GD, Sequeira E, Sherry ST, Sirotkin K, Souvorov A, Starchenko G, Tatusov RL, Tatusova TA, Wagner L, Yaschenko E.(2007) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*. **35**: D5-12.

P-29

Transcriptome wide analysis for lncRNAs identification of primary and secondary hair follicles of goat (*Capra hircus)*

Himansu Kumar[1], Jaya Pandey[1], Sarika Sahu[1], Nazir A Ganai[2], A. R. Rao[1]

Centre for Agricultural Bioinformatics, ICAR-Indian Agricultural Statistics Research Institute, Pusa, Library Avenue, New Delhi- 110 012, India, Division of Biotechnology, Sher-e-Kashmir University of Agricultural Science and Technology of Kashmir, India Large scale transcriptome

analysis of different organisms revealed that most of the genomic components are non expressive and referred as non coding RNAs. Consisting of miRNAs, siRNAs, piRNAs, snoRNAs etc. Long non coding RNAs (lncRNA) are longer than 200 nucleotides and play a significant role in gene regulations of goat. However, exploration of lncRNAs in the context of primary and secondary hair follicles is limited. In this study, genome sequence of *Capra hircus* has been retrieved from Goat Genome Consortium and considered as reference genome whereas transcriptome sequences of the primary hair follicles (PHF) and secondary hair follicles (SHF) of *capra hircus* goat are taken from public domain. Reference genome of the *Capra hircus* has been indexed with the help of Bowtie2, and the transcriptome sequences of the *Capra hircus* have been trimmed for the purpose of quality check. TopHat aligner was used to align the trimmed sequences against the indexed reference genome and sequences having less than 200 nucleotides have been excluded. The selected sequences have been aligned with the swissprot database to filter out the potential non protein coding RNAs. Out of 402456 PHF and 68784733 SHF sequences, a total of 3000 PHF and 8108 SHF were filtered out as likely long non coding RNAs. These were subjected to the estimation of coding potential by running Coding Potential Calculator (CPC) to infer the quality and completeness of lncRNA on the basis of protein coding score. A total of 439 and 177 sequences of PHF and SHF respectively were finally filtered as lncRNAs. The tRNAs, snRNAs, and snoRNAs were then eliminated from the filtered lncRNAs by scanning through their respective available databases and the rest of the sequences were considered as predicted lncRNAs. The Blastn program was used to remove the probable miRNAs and UTR sequences. The selected lncRNAs were further analyzed for their function as well as prospective role in gene regulations in goat.

P- 30

Immunoinformatic identification of potential epitopes against Salmonellosis

Drashya Sharma[# a,b], Bhrugu Yagnik [a,b], Harish Padh[b], Priti Desai*[a]

[a]Department of Cell and Molecular Biology, B. V. Patel Pharmaceutical Education and Research; Development Centre, Ahmedabad, India-380054; [b] Sardar Patel University, Vallabh Vidyanagar, Gujarat, India-388120
E-mail: [#]sharma.drashya@gmail.com, bjyajnik3@gmail.com, hpadh@yahoo.com, [*]preetindesai79@gmail.com

Infection through *Salmonella* is still the major cause of morbidity and mortality. Antibiotic treatment is not much effective as several strains of *Salmonella* have developed resistance to multiple antibiotics(1). Therefore, vaccine development would be strategic alternative to prevent the infection. Previous efforts of vaccine development were not completely successful in the effective *Salmonella* infection management (2). The development of immunoinformatics has aided the rational development of protective prophylactics (3). Hence, the present study immunoinformatically identified potential epitopes against Salmonellosis. A sequential workflow was followed to identify potentially immunogenic epitopes. These predicted epitopes were derived from outer membrane proteins and Flagellin, conserved among several *Salmonella spp*.and contribute in the pathogenesis of *Salmonella*.

In the present work, B cell and T cell recognised epitopes which are derived from Outer membrane protein F (OmpF), Outer membrane protein 28 (Omp 28) and Flagellin protein (FliC) were predicted. The B cell epitopes were found to be topologically present on surface exposed loops of OmpF and Omp28 and outer exposed â sheet region (D2 and D3) of FliC proteins. Surface exposed regions are the potential site for antibodies binding, indicating the humoral immunogenic potential of selected epitopes. Additionally, T cell epitopes for both the MHC class I and II were predicted. In order to have a broader coverage, the epitopes displaying binding to maximum number of MHC class I and II alleles were selected. The epitopes were analysed for Indian and global population coverage and are expected to be effective against the large population of the world. In conclusion, we generated the pool

of potent B cell and T cell recognised epitopes capable of eliciting humoral and cellular arms of immunity against *Salmonella*. Future experimental validation of predicted epitopes will ensure the immunogenicity and could be considered for developing a multivalent candidate for vaccine development against Salmonellosis.

References

[1] Taylor DN, Gardner P, Nichols RL, Ramsey KM, Smith LG, Warren JW. 2003. Infectious Diseases VI.

[2] Garmory HS, Brown KA, Titball RW. 2002. Salmonella vaccines for use in humans: Present and future perspectives. FEMS Microbiol Rev 26:339–353.

[3] Sharma D, Patel S, Padh H, Desai P. 2016. Immunoinformatic Identification of Potential Epitopes Against Shigellosis. Int J Pept Res Ther 22:481–495.

P-31

Elucidating thesoft spot CYP3A4 metabolic profiling of small molecule kinase inhibitors

Chelli Sai Manohar[a#], B. Siva Kumar[a]

[a]RIMM Lab, Department of Chemistry, Sri Sathya Sai Institute of Higher Learning
[#]E-mail: chellisaimanohar@sssihl.edu.in, bsivakumar@sssihl.edu.in

Given the tedious and expensive experimentation in the soft spot metabolic analysis of drugs, we propose a cost effective and reliable *in silico* strategy of metabolic profiling. In this context, we firstly compared the structures of the metabolizing class of enzymes and the corresponding drugs metabolized. Further, the thermodynamics of this spontaneity of metabolism in 95 approved drug molecules obtained from the drug bank was evaluated. Finally, we confirmed the reaction mechanism via computational simulations and docking studies. In this regard, the *ab initio* DFT studies delineated the energy profile for the CYP metabolic pathway for a sample kinase inhibitor, while the docking studies for 14 kinase inhibitors ascertained the explicit role of Fe ion in this process. The reliability of this strategy is confirmed by comparison to the reported experimental metabolites for the corresponding drugs. Thus we arrive at a reliable *in silico* strategy for the early prediction of the soft spots in drugs labile to metabolism to model the bio-availability prior to synthesis in drug design.

P- 32

Improved recognition of heat shock proteins, their families and sub-types based on *g*-spaced di-peptide features and support vector machine

Shachi Gahoi[1#*], Prabina Kumar Meher[2], Tanmaya Kumar Sahu[1] and Atmakuri Ramakrishna Rao[1]

[1]Centre for Agricultural Bioinformatics, ICAR-Indian Agricultural Statistics Research Institute, New Delhi, India; [2]Division of Statistical Genetics, ICAR-Indian Agricultural Statistics Research Institute, New Delhi, India

Heat shock proteins (HSPs) are important ingredients for cell growth and variability. Since conventional approaches are expensive and large number of protein sequences has been generated in the post-genomic age, it is highly desired to develop an automatic and accurate computational method for predicting HSPs, their families and sub-types. Thus, we developed a computational approach using G-spaced di-peptide compositions as input features and SVM as the prediction machine. The proposed approach achieved an overall accuracy of ~84% in predicting HSPs, ~97% in predicting six different families of HSPs and ~94% in predicting four types of DnaJ proteins, with bench mark datasets. The proposed method was further evaluated for proteome-wide prediction of HSPs by using proteome datasets of eight different species. The result revealed that ~50% of the predicted HSPs in each species have HSP domains using InterProScan. For easy prediction of HSPs, a user friendly online prediction server is being developed. The developed approach is believed to supplement the existing methods for prediction of HSPs, their families and sub-types.

P- 33

In silico Molecular Docking Analysis and Simulation Studies on Apicoplast Proteins of Malaria Parasite

Drista Sharma[#] and Tarun Kumar Bhatt[*]

Department of Biotechnology, School of Life sciences, Central University of Rajasthan, Ajmer, India
E-mail: drishta.sharma113@gmail.com, tarun@curaj.ac.in

Drug-resistant malaria raises a cardinal public health problem and new antimalarial drugs are a great necessity. The breakthrough of a vestigial plastid organelle has intensified an escalation of studies in search of antimalarial drugs. The role of apicoplast in the life of the parasite is peerless. The housing of different metabolic processes which are essential for parasite survival adds onto the indispensability of the apicoplast. In the present investigation, a series of compounds from different small molecule libraries (ZINC, NCI, etc) and some chemically synthesized compounds having therapeutic properties were evaluated for their antimalarial activity using combined *in silico* and *in vitro* approach against three apicoplast specific targets, namely caseinolytic protease, peptide deformylase and oxoacyl ACP reductase. The binding energy was assessed using Glide program. The conformers showing significant binding score and interactions with the key residues of targets were prioritized for the exploration of their natural dynamic behavior in solution on a different timescale. The dynamic characteristics of the docked complex were investigated in aqueous solution for 100ns by molecular dynamics simulation. The position fluctuations of different hits were explored. The plateau phase with respect to RMSD, RMSF plots and invariant Rg values revealed the efficacious nature of interactions and stability of the complex. The protein-ligand interaction in the aqueous environment indicated the presence of strong hydrogen bond with significant occupancy. On the basis of binding scores and stability, seven (in the case of caseinolytic protease and peptide deformylase) and five hits (in case of oxoacyl ACP reductase) were recognized as potential candidate hits. The inhibitory activity of putative drug analogue was estimated against P. falciparum 3D7 by *in vitro* parasite inhibition assay. The results were in concordance with the efficacy of conformers. The study is expected in the disclosure of some inhibitory molecules with the view of developing them into antimalarials against indispensable organelle, apicoplast.

P- 34

Bioinformatic Tools & Databses: Celiac Disease

Shivani Rustagi[1#*], Chanu Handa[1]

[1]Amity Institute of Food Technology, Amity University, Noida, Uttar Pradesh, India – 201313
E-mail: rshivani.foodtech@gmail.com, chanda@amity.edu

Around the globe about 8% children and 5% adults are afflicted with food allergies against milk, eggs or wheat. Celiac Disease (CD) is a T-cell mediated autoimmune disease of the gastrointestinal system characterized by intake of gluten from wheat, barley, rye. Currently, the only effective treatment is lifelong adherence to gluten for CD patients. To comprehend the nature and implications of allergies, various new methods like genomics, proteomics, transcriptomics, metabolomics and microarray have been developed. The resulted extensive information is incorporated into a logical set of data to give discrete public databases. The Food Allergy Research & Resource Program (FARRP), ProPepper and GluPro V1.0 are such specialized database encompassing distinctive series of known and putative allergens distilled from NCBI GenBank, EMBL, PIR & SWISS-PROT. These databases also offer different tools for further analysis of food allergens. To reduce the risk of exposure of gluten-sensitive patients, proteins should be assessed regarding their ability to elicit CD. Proficient screening tools have been given by the CD peptide and protein searchable database of FARRP. They have enlisted two methods: Exact peptide match algorithm and a FASTA algorithm to identify a protein representing a risk of eliciting CD (Allergen Online, 2017). ProPepper contains datasets that are collected from several open databases, manually curated, annotated and interpreted in three main data tables: Protein-, Peptide- and Epitope list views (Juhász *et al.,* 2015). GluPro V1.0 is first manually curated public-source wheat gluten protein sequence database (GluPro V1.0) in a FASTA format to support the application of proteomic methods for gluten

protein detection and quantification (Juhász *et al.*, 2017). Bioinformatics and knowledge integration have been playing facilitating role in the science of food allergies. Bioinformatics is an integral component of any research activity being popular due to its ability to analyse huge amount of data quickly and cost effectively.

References

[1] Allergen Online. 2017. *Celiac Disease (CD) Novel Protein Risk Assessment Tool*. University of Nebraska-Lincoln accessed on 10 July 2017. <*http://www.allergenonline.org/celiachome.shtml*>)

[2] Bromilow S, Gethings LA, Buckley M, Bromley M, Shewry PR, Langridge JI and Mills EC. A curated gluten protein sequence database to support development of proteomics methods for determination of gluten in gluten-free foods. Journal of Proteomics. 2017; 163: 67-75.

[3] Juhász A, Haraszi R and Maulis C. ProPepper: a curated database for identification and analysis of peptide and immune-responsive epitope composition of cereal grain protein families. Database. 2015.

P-35

PeMtb: a database of MHC antigenic peptide of *Mycobacterium tuberculosis*

Ziaul Hasan[1,#*], Qamar Zia [2,3#] , Asim Azhar [3,#], Shadab Ahmad [3,#] , Mohammad Afsar [4], Mohammad Owais [3], Mahfooz Alam [4], Shabab Akbar [4], Ghulam Md Ashraf [5], Swaleha Zubair [6], Gjumrakch Aliev [7,8,9]

[1]Department of Biosciences, Jamia Millia Islamia, New Delhi, India; [2]Department of Biotechnology, Gagan College of Management and Technology, Aligarh, India; [3]Interdisciplinary Biotechnology Unit (IBU), Aligarh Muslim University, Aligarh, India; [4]Glocal Agro-Med Informatics Research Institute (GAMIRI), New Delhi, India; [5]King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia; [6]Women's College, Aligarh Muslim University, Aligarh, India; [7]GALLY International Biomedical, Research Consulting LLC., 7733 Louis Pasteur Drive, #330, San Antonio, TX, 78229, USA; [8] School of Health Science and Healthcare Administration, University of Atlanta, E. Johns Crossing, #175, Johns Creek, GA, 30097, USA; [9] Institute of Physiologically Active Compounds Russian Academy of Sciences, Chernogolovka, 142432, Russia

*# These authors contributed equally, *Presenting author*

E-mail: zhasan.biochem@gmail.com

In order to design a subunit vaccine for tuberculosis, identification of antigenic T-cell epitope is of utmost importance. Several MHC prediction server are available that can accurately predict antigenic peptide of variable lengths. Ironically, peptides predicted from one server not necessarily fulfill the criteria projected by another server, thus creating a confusing situation for scientists to choose a best epitope. Keeping this problem in mind, we developed a comprehensive database of peptides based on various immunogenic Mycobacterial proteins that had been already ranked by more than one existing MHC prediction servers. For each protein, PeMtb generates a set of three different mers of variable lengths (12 mer or 13-mer) based on their ranking; with each mer being predicted for a plethora of MHC alleles. The database offers selection of the peptide (mers) with best binding affinity as predicted by several available servers. The database provides quick information; thus cut short efforts that were required earlier in searching MHC prediction servers. All the resources and information can be accessed by PeMtb home page *www.pemtb-amu.org*. We trust and anticipate that PeMtb will be a practical platform for trial and computational analyses of antigenic peptides for the pathogen *Mycobacterium tuberculosis* and will be helpful in designing of an effective subunit vaccine against tuberculosis

P- 36

Relative significance of genomic traits directs expression level and breadth

Manju Bansal and Sanjukta Das

Molecular Biophysics Unit, Indian Institute of Science, Bangalore

Understanding the structural organisation of genome and various gene properties has helped in elucidating gene regulation and expression at cellular and molecular levels. In higher eukaryotes, gene structure and promoter architecture have emerged as significant factors influencing variation in number of transcripts (expression level) and specificity of gene expression in a tissue (expression breadth), which eventually shape the phenotype. In this study, transcriptome data of different tissue types at various developmental stages of *A. thaliana*, *O. sativa*, *S. bicolor* and *Z. mays* have been used to understand the relationship between properties of

gene components and its expression. Our findings revealed that in plants gene architecture as well as promoter properties are significantly linked to gene expression level as well as breadth, suggesting interconnected molecular and functional pathways. Interestingly, results from multiple regression analysis, which predicts a significant amount of change in expression level and breadth is influenced by a particular genomic trait, revealed that intron content of primary transcript (as %) is a powerful determinant of tissue specificity. Similarly, among structural properties of the promoter, stability was found to be negatively linked to expression breadth, while DNase1 sensitivity strongly governed gene expression breadth in monocots and gene expression level in dicots. This differential response to DNase1 sensitivity exhibited by different plants led us to examine the compositional distinction in their promoter regions, and further analysis revealed that tissue specific genes are enriched with TATA box and Y-patch. An extensive study on the size of the orthologous groups and gene expression parameters illustrated that multi-copy orthologous genes in plants are longer, highly regulated and tissue specific.

P- 37

Functional annotation and identification of putative drug target in Vaccinia Virus

[1]Yashbir Singh; [2]Nemi Malhotra; [3]Himanshu Narayan Singh, *Satyvani Guttula

[1]Indian Institute of Information Technology, Allahabad; [2]Jamia Hamdard University, Mehrauli Badarpur, New Delhi; [3]All India institute of Medical Science, New Delhi
**Indian Institute of Information Technology, Allahabad*
Email: yashbir143@gmail.com, nemi.malhotra@gmail.com, himanshu720@gmail.com, guttula.satyavani@gmail.com

Introduction: *Vaccinia virus* (VV) has proved a great help to the mankind for eradicating life threatening pox disease. In past two decades it has been realised that its role was not limited to develop vaccine only but extended widely as potential therapeutics against cancer and understanding several mechanisms related to virus host interactions. With the advent of new technologies in field of molecular biology and immune-therapeutics, this virus has taken a prime and promising place to employ as therapeutic agent against cancer. Oncolytic potential of *Vaccinia virus* has been documented significantly through early clinical trials.

Vaccinia virus has been used more broadly for human immunization than other vaccine. In this analysis, *Vaccinia virus* data exhibited that 45% proteins are conserved hypothetical proteins whose function is not determined yet. We used some software and some computational approach for the functional annotation and identification of HPs for the genome of Vaccinia virus. We annotated functions of the hypothetical protein using machine learning approach.

Methodology: Analysis of 100 HPs from *Vaccinia virus* was done using novel computational approach. We precisely predicted the functions of 45 HPs with sub-cellular localization, physicochemical parameters prediction and the ROC analysis successfully characterized the systematic performance assessments. We considered only HPs annotated with the high level of confidence. This result shows functional importance of HPs in the survival of the pathogen in the host.

Conclusion: Our study facilitate to identify the hidden function of HPs and disorders which is potential therapeutic targets, may play a role in better understanding of host- pathogen interactions and further research for new inhibitors and vaccine can be conducted.

References

[1] Naqvi, A. A. T., Ahmad, F., & Hassan, M. I. (2015). Identification of functional candidates amongst hypothetical proteins of Mycobacterium leprae Br4923, a causative agent of leprosy. *Genome*, *58*(1), 25-42.

P- 38

Mutational analysis of microbial ferulic acid decarboxylase towards enhancement of binding affinity: an *in-silico* approach

Pravin Kumar, Shaswati Ghosh Sachan, Raju Poddar

Department of Bio-Engineering, Birla Institute of Technology-Mesra, Ranchi, JH 835 215, India.

Microbial Ferulic Acid Decarboxylase (FADase) catalyses the conversion of ferulic acid to 4-

hydroxy-3-methoxystyrene (4-vinylguaiacol) via non-oxidative decarboxylation. We present computational three dimensional structural and functional analysis of FADase from Enteroter sp. Px6-4 (3NX1) for better bindings of substrate. Enzymatic catalytic site and binding site are critically explored. Sequential site directed mutation of those sites are introduced. Active sites of mutated FADases (SNPs) are analyzed with dynamic cross-correlation map (DCCM) and principle components analysis (PCA) method. Further, structures were validated and optimized through energy minimization. Docking studies were carried out between ferulic acid and different SNPs. It was observed that, certain amino acids like Tyr21, Trp25, Tyr27 and Glu134 at active sites are responsible for proton transfer to substrate. Further, it was observed that mutated form Y27F (Tyr27Phe) of FADase shows better binding affinity towards ferulic acid than its native form.

P- 39

Proteus: A Random Forest Classifier to Predict Disorder-to-Order Transitioning Binding Regions in Intrinsically Disordered Proteins

Sankar Basu[#1,2*], Fredrik Söderquist[1], Björn Wallner[1*]

[1]Bioinformatics Division, Linköping University, Linköping, Sweden; [2]Department of Chemistry, University of Delhi, Delhi – 110009, India
[#] Presenting Author
**Corresponding Author: E-mail: nemo8130@gmail.com*

The focus of the computational structural biology community has taken a dramatic shift over the past one-and-a-half decade from the classical protein structure prediction problem to the possible understanding of intrinsically disordered proteins (IDP). The current interest lies in the unraveling of a disorder-to-order transitioning code embedded in the amino acid sequences of IDPs. Disordered proteins are characterized by enormous amount of structural plasticity which makes them promiscuous in binding, multi-functional in cellular activity and atypical in folding energy landscapes resembling partially folded molten globules. Also, they are attractive drug targets due to their involvement in several human diseases. The study of the structural ensemble of IDPs is rather difficult for transient interactions. When bound to a structured partner, an IDPR adapt an ordered conformation. The residues that undergo this disorder-to-order transition are called 'protean' residues, and the first step in understanding the interaction involving an IDPR would be to predict these 'transitioning' interacting residues. There are a few available methods which predict these protean segments from their amino acid sequences, however, their performance reported in the literature leaves clear room for improvement. In this background, the current study presents 'Proteus' [1], a random-forest-based protean predictor that predicts the likelihood of a residue to be 'protean'. Proteus compares favorably with existing methods predicting twice as many true positives as the second best method (55% vs. 27%) at a much higher precision on an independent data set. The current study also sheds some light on a possible 'disorder-to-order' transitioning consensus, untangled, yet embedded in the amino acid sequence of IDPs.

Reference

[1] Proteus: A Random Forest Classifier that Predicts Disorder-to-Order Transitioning Binding Regions in Intrinsically Disordered Proteins, Sankar Basu, Fredrik Söderqvist, Björn Wallner*

[2] Journal of Computer-Aided Molecular Design, 2017, Volume 31, Issue 5, pp 453–466DOI:10.1007/s10822-017-0020-y.

P- 40

A Robust Algorithm for Measuring Surface Curvature of Protein

Abhijit Gupta[#], Arnab Mukherjee[*]

Department of Chemistry, IISER, Pune
E-mail: abhijit.gupta@students.iiserpune.ac.in, arnab.mukherjee@iiserpune.ac.in

Protein surface shape and curvature are key aspects of protein function and recognition. Structural information of protein surface regions would be helpful in detailed studies of the relationship of protein structure and function. The surface of protein is not uniform, and hence it's not amenable to differential geometry approach, which requires a smooth, differentiable surface. Our methodology

involves division of the protein surface into patches and fitting a sphere to each patch. The radius of the fitted sphere could then be used for calculating surface curvature (ê=1/r), where *r* is the radius of the fitted sphere.

We have employed hierarchical clustering with farthest neighbour approach and Ward's method as clustering criteria[1] to divide protein surface into patches. It takes into account the nuances in surface topology, unlike previously available approaches that used either solid-angle approach for curvature measurement or pick points within a local radius to deûne the patch for curvature measurement.[2]

For accurately fitting sphere to each patch, we have extended the 'Hyper accurate' algebraic fit algorithm for circle fitting[3] into 3D and formulated a non-iterative algorithm for least square fitting of sphere to a surface.The patch size can be adjusted by varying threshold value in clustering criterion to highlight atomic features or larger features such as potential peptide binding sites. We have employed our methodology to study effect of binding of ligand to a protein on surface curvature, variation of charge density with curvature and protein-protein interaction.

Further, we are looking forward to classifying different classes of proteins based on the surface curvature criterion.

References

[1] Matej Francetiè, M.Nagode,B.Nastav, Hierarchical Clustering with Concave Datasets, Metodološki zvezki, Vol. 2,No.2,2005,173-193.

[2] Coleman, Burr, M.A., Souvaine, Cheng, (2005), An intuitive approach to measuring protein surface curvature. Proteins, 61: 1068–1074. doi:10.1002/prot. 20680.

[3] A.Al-Sharadqah, N.Chernov, Error Analysis of Circle Fitting algorithms, arXiv:0907.0421

P- 41

In silico Analysis of the Distribution of Simple Sequence Repeats in Plastic Eating Bacterium *Ideonella sakaiensis*

Srinija Bondada[1], Kranthi Chennamsetti[1] and Suresh B. Mudunuri[1#*]

[1]Centre for Bioinformatics Research (CBR), SRKR Engineering College, Chinna Amiram, Bhimavaram, Andhra Pradesh, India - 534 204

** sureshverma@gmail.com*

Plastic usage has been increased exponentially during the last few decades and more than 8 million tons of plastic are being dumped in our oceans every year. A recently sequenced novel bacterium *Ideonella sakaiensis* has a peculiar characteristic of consuming plastic waste and metabolizing it for its growth [1]. This is the only strain of class betaproteobacteria that uses plastic as its primary carbon and energy source by hydrolysing PET (polyethylene terephthalate) and can be a potential solution for biological degradation of plastic waste. As it is recently sequenced, not very much is known about its evolution and dynamics. Analyzing the Simple Sequence Repeats (SSRs), especially in genic regions, would throw some light about the evolution of this organism. SSRs, also known as Microsatellites, are short tandem repeats of size 1-6 bp, are known to induce plasticity in bacteria & viruses and are also used as potential markers. So, we have performed a short survey of the distribution of SSRs in genic regions of this bacterium. Whole genome shot-gun sequences of *Ideonella sakaiensis 201-F6* strain have been obtained from NCBI and the corresponding gene sequences have been extracted. A total of 5,375 genes have been processed with Imperfect Microsatellite Extractor (IMEx) [2] tool for identifying imperfect repeats (imperfection of 10%) with repeat numbers: Mono-10, Di-5, Tri-4, Tetra-3, Penta-2, Hexa-2. It is observed that trinucleotide repeats are found to be abundant, followed by penta, hexa, tetra and dinucleotide repeats. Surprisingly, no mononucleotide repeats are found in the gene regions. On an average, each gene has around 5-6 SSRs indicating their positive role in its evolution. When tri-nucleotide repeats are considered, the amino acid Alanine is found to be encoded predominantly. A detailed report of frequent SSRs

and their distribution generated in this study can form a very useful resource.

References

[1] Tanasupawat, Somboon, et al. "Ideonella sakaiensis sp. nov., isolated from a microbial consortium that degrades poly (ethylene terephthalate)." *IJSEM*, 66.8 (2016): 2813-2818.

[2] Mudunuri, Suresh B., and Hampapathalu A. Nagarajaram. "IMEx: Imperfect Microsatellite extractor." *Bioinformatics* 23.10 (2007): 1181-1187.

P- 42

Riboswitches in Archaea Phylogenetic analysis and the sub-family specific conservation patterns of SWEET sugar transporter family in eukaryotes and prokaryotes

Ankita Gupta[#] and Ramasubbu Sankararamakrishnan[*]

Department of Biological Sciences and Bioengineering Indian Institute of Technology Kanpur E-mail: gankita@iitk.ac.in, rsankar@iitk.ac.in

SWEET (Sugar Will be Eventually Exported Transporter) sugar transporter is a recently identified family of integral membrane proteins [1]. They are present in both eukaryotes and prokaryotes and play a crucial role in different physiological processes. Their prominent role in phloem loading, pollen nutrition, seed filling and pathogen susceptibility in major crop plants have made them important targets for sustainable food production. Till date, three crystal structures from prokaryotes and one eukaryotic SWEET structure have been elucidated. The functional pore of this uniporter is composed of two SWEET/Mtn3 domains. In eukaryotes, the SWEET proteins have 7-transmembrane helices with an internal structural symmetry, where the two triple-helix-bundles (SWEET domains) are fused with the help of a linker transmembrane helix 4. In prokaryotes, SWEET homologs contain single triple-helix-bundles (Semi-SWEET), which dimerize to form a functional transporter [2]. We have performed exhaustive structure-based sequence analysis of plant (979 sequences), metazoan (324 sequences), bacteria (728 sequences) and archaea (48 sequences) SWEET proteins, and have generated homology models using the crystal structures as templates. Subsequent phylogenetic analysis reveals SWEET subfamilies present in higher organisms. We found 16 positions conserved across all the SWEET proteins, out of which 7 are pore-facing, and the rest occur at helix-helix interfaces or at the transporter gates. Other pore-facing positions show kingdom or sub-family specific conservation patterns. We have also identified conserved positions which might be crucial for higher oligomerization states of SWEET proteins, present in eukaryotes. This study highlights key features of SWEET transporters which distinguish between different types of sugars, and the residues playing role in transporter function and structural stability. Molecular features distinguishing SWEET proteins of bacteria, archaea, and eukaryotes to investigate the evolution of SWEETs in eukaryotes from Semi-SWEET in prokaryotes are currently in progress.

References

[1] Chen *et al*. Nature 468,527–532(2010).

[2] Tao *et al*. Nature 527,259-63(2015).

P- 43

Overlapping Brain Disorders in Human Dopamine Receptors Interactions Network

Avijit Podder[#] and N. Latha*

*Bioinformatics Infrastructure Facility, Sri Venkateswara College (University of Delhi), Benito Juarez Road, Dhaula Kuan, New Delhi 110021, India, *Email: lata-bic@svc.ac.in*

Dopamine receptors (DRs) are the members of G protein-coupled receptor (GPCR) and widely expressed in the central nervous system. Intercommunication of DRs with their associate protein partners is crucial to maintain regular brain function in human. Majority of the brain disorders arise due to malfunctioning of such communication process. Hence, contributions of genetic factors, as well as phenotypic indications for various neurological and psychiatric disorders are often attributed as sharing in nature. The mechanism of the complex brain disorders cannot be interpreted by a single gene product or the behaviour of a specific individual pathway. A holistic approach is required to interpret the disease phenotype.

Comprehensive interactions of candidate proteins associated with human dopamine receptors were captured through protein-protein interaction network (PPIN) analysis. A holistic map of human Dopamine Receptors Interactions Network (DRIN) was constructed for all five DRs and their associate protein partners by mapping them into human interactome [1]. Here, we have investigated our studied network – DRIN to identify the genetic overlapping in various neurological disorders. We applied a series of statistical and computational approaches to understand the sharing nature of disease genes in dopamine receptors connecting protein-protein interactions network. Our study pinpointed certain common drug targets those are valuable in treatment of overlapping brain disorders in the network [2].

References

[1] Avijit Podder, Nidhi Jatana, N. Latha (2014). Human Dopamine Receptors Interaction Network (DRIN): A systems biology perspective on topology, stability and functionality of the network. Journal of Theoretical Biology, 357:169-183.

[2] Avijit Podder and N. Latha (2017) Data on Overlapping Brain Disorders and Emerging Drug Targets in Human Dopamine Receptors Interactions Network. Data in Brief. 12:277–286.

P- 44

Metagenomics Analysis of Microbial Community in Riverine Ecosystem in Northern India

Indra Singh[1], Biswanath Patra[2], H. J. Chakraborty[2], B. K. Behera[2] and A. R. Rao[1]

[1]ICAR-Indian Agricultural Statistics Research Institute, New Delhi-110012
[2]ICAR-Central Inland Fisheries Research Institute, Barrackpore, Kolkata-700120

Microbes are present in the riverine ecosystem of Northern India, where they play important role to stabilize the ecosystem. Water quality of the river has deteriorated day by day due to anthropogenic activities, population growth and rapid industrialization. There are various studies on river pollution but very limited information is available on metagenomics of the microbiota of rivers like Ganga. Therefore, the present work has been designed to study the microbiota of the River using metagenomics for better understanding of distribution of microbial diversity, antimicrobial-resistance and presence of virulence-markers. Therefore, soil / sediment and water samples were collected for metagenomics studies at different sites of Ganga. The metagenomics profiling has been made for polluted sites verses control (fresh water) using tools like, Prinseq, Velvetg, Metavelvet and Phylopathia. From the present study microbiota that are helpful or harmful for the host or environment have been evaluated, compared and identified. Analysis of data and binning has revealed that there is a disproportionate representation of various metabolism signalling pathways present in polluted sediment samples as compared to those found in control / freshwater samples.

P- 45

Investigation of Host-Pathogen Interaction Interface at the early HIV-1 infection stage using computational approaches

Mansi Pandit#, Deeksha Pandey, Avijit Podder and N. Latha*

Bioinformatics Infrastructure Facility, Sri Venkateswara College (University of Delhi), Benito Juarez Road, Dhaula Kuan, New Delhi 110021, India
**Email: lata-bic@svc.ac.in*

Human Immunodeficiency Virus-1 (HIV-1) is the major causative agent for Acquired Immune Deficiency Syndrome (AIDS). The enveloped virus initiates infection by attachment of its envelope glycoprotein gp120 with host cellular receptor CD4. This interaction triggers the activation of an array of proteins from both viral and human proteomes which play important roles during host-pathogen interaction. Apart from the structural glycoproteins (gp120 and gp41), several regulatory (Tat and Rev) and accessory (Vpu, Vpr, Vif and Nef) proteins from the virus are also a part of the early phase of infection. Hence, the study of proteins involved in early infection becomes crucial to obtain a comprehensive understanding of the disease mechanism. The lack of crystal structure of some of these proteins (gp41, Nef and Tat) provided us the scope to use computational protein modeling

techniques and predict their three-dimensional models. The predicted protein structures were then refined using molecular dynamics simulations. Besides, molecular docking was performed to determine the desirability of these target proteins for already available HIV-1 specific drugs which indicates the usefulness of these protein structures to identify effective combination drug therapy for AIDS. The study indicated towards a unique combination of two marketed HIV-1 specific drugs Delavirdine and Indinavir for potential use in combination antiretroviral treatment. The significant interaction of these drugs with gp 41, Nef and Tat proteins also highlighted drug scaffolds that can be used as seed for design of target specific inhibitors.

References

[1] Deeksha Pandey, Avijit Podder, Mansi Pandit & N. Latha*. (2016). CD4-gp120 Interaction Interface - A Gateway for HIV-1 Infection in Human: Molecular Network, Modeling and Docking Studies. *Journal of Biomolecular Structure and Dynamics.* 29:1-14

P- 46

Molecular dynamics simulation of Burkholderia cepacia lipase in organic solvents

A. C. Mathpati# and B. M. Bhanage*

Department of Chemistry, Institute of Chemical Technology, Mumbai 400 019
*Email: *bm.bhanage@ictmumbai.edu.in, #acmathpati@gmail.com*

Lipases, a subclass of hydrolases, have gained a lot of importance as they can catalyze esterification, transesterification and hydrolysis reaction in non-aqueous media. Solvents affect the enzymes or enzyme"substrate complexes by producing changes in the conformational rigidity of enzymes, the active site, or altering the solvation of the transition state. The activity of lipases strongly depend on the logP value of solvents. Molecular dynamics and docking studies can help to understand the effect of solvents on lipase conformation and free energy changes in molecular transformations [1-2]. In this work, molecular dynamics (MD) simulations of *Burkholderia cepacia* lipase (BCL) using gromacs have been carried out for 50 ns for solvents such as acetonitrile (logP -0.15), diethyl ether (logP 0.8), toluene (logP 2.7) and hexane (logP 3.5). The NVT simulations were carried out for 200 ps, followed by NPT simulation for 500 ps for equilibration. Energy minimisation was carried out using the steepest descent method. Root mean square deviation (RMSD), radius of protein gyration and B factor were studied to evaluate stability of conformations along MD simulation. The RMSD values were within the range of 0.15 to 0.20 nm and radius of gyration was found to be with 1.65 to 1.9 nm. The distribution of B factors along a protein sequence is regarded as an important indicator of the protein's structure, reflecting its flexibility and dynamics. Major changes in the B factor compared to reference structure were observed between residues 60 to 80, 120 to 150 and 240 to 260. Higher unfolding was observed in toluene and diethyl ether compared to hexane and acetonitrile. The catalytic triad of BCL consist of Ser87, Asp264 and His286 residues. The critical bond distances have been estimated based on the MD simulations.

References

[1] Trodler, P., Schmid R.D., and J. Pleiss (2009). Modeling of solvent-dependent conformational transitions in Burkholderia cepacia lipase. BMC Structural Biology, *9*:38, DOI:10.1186/1472-6807-9-38.

[2] Mathpati A.C. and Bhanage B.M. (2016), Combined docking and molecular dynamics study of lipase catalyzed kinetic resolution of 1-phenylethanol in organic solvents. Journal of Molecular Catalysis B: Enzymatic. DOI: 10.1016/j.molcatb.2016.12.005

P- 47

Biological Networks in the study of Neurological Disorders

Abhinav Pokhriyal, Aishwarya V., Anushka Takhi, Bhoomika Manchandia, Mehar Monga#, Prerna Sabharwal, Radhika Sinha#, RohanKapoor, Simarjot Kaur, Tanushka Rana# and N. Latha*

Sri Venkateswara College (University of Delhi)
Benito Juarez Road, Dhaula Kuan, New Delhi- 110021, India
**Email: lata@bic-svc.ac.in*

Neurological disorders pose a major public health issue that affects billions globally. There is an unmet

medical need for the development of novel medicines for neurological disorders, as novel treatment approaches have stalled for decades. Developing new drugs are more challenging than the conventional central nervous system (CNS) drugs for such complex polygenic disorders. Therefore, a holistic approach is indispensible for the therapeutic intervention of such debilitatingdiseases. In our study, comprehensive interactions network of candidate proteins implicated in three neurological disorders; namely: Alzheimer's, Bipolar and Parkinson's disease were constructed and analyzed to gain insight into the disease etiology. In this study, we have analyzed various topological parameters such as degree, betweenness, closeness and clustering coefficient of the constructed network. Our findings would help to prioritize diseased proteins in the network that may have the potential to act as candidate drug targets in future drug discovery endeavors.

P- 48

Evolutionary patterns in prion proteins and its incidence in India

Sonia Mukherjee[#], Pratim Chakraborti[*]

Drug Repurposing, Excelra Knowledge Solutions Pvt Ltd
E-mail: [#]sonia.jejo@gmail.com, [*]protimster@gmail.com

Prion proteins infect mainly mammals and manipulate the normal functioning of cellular proteins. The infected cellular proteins fold abnormally leading to debilitating neurological diseases in humans and animals. Prion diseases can arise sporadically, be inherited, or acquired through infected food, surgical instruments and iatrogenic means i.e., through blood transfusion. Although the first recorded incidence of prion disease in man dates back to 1920, established protocols for disease diagnosis and effective treatment regimen are yet to be formulated. Surprisingly, there is little or no information about the evolutionary relationship of the prion proteins in mammals. This study focuses on the evolution of prion proteins by employ ingiterative PSI BLAST. The prediction tool compares the human prion protein to the prions found in distantly related species thus increasing the coverage of protein members of a family and finds a relationship between the superfamilies' of the protein. The prion protein phylogeny was traced which indicated that the prion protein coevolved with its host. A part of this study also ties to consolidate the epidemiological evidences of prion disease in a few hospitals in India as well as explore its relation with the food habits of affected individuals.

References

[1] Prusiner SB: Prions. ProcNatlAcadSci USA 1998, 95:13363-13383. Smirnovas V, Baron GS, Offerdahl DK, Raymond GJ, Caughey B, Surewicz WK: Structural organization of brain-derived mammalian prions examined by hydrogen-deuterium exchange. Nat StructMolBiol 2011, 18:504-506.

[2] Imran M, Mahmood S: An overview of animal prion diseases. Virol J 2011, 8:493.

[3] Collinge J, Clarke AR: A general model of prion strains and their pathogenicity. Science 2007, 318:930-936.

P- 49

An in silico toxicity assessment of DEHP and its metabolites on metabolism of essential amino acid

Neha Singh[#], Vikram Dalal, Pravindra Kumar*

Department of Biotechnology, Indian Institute of Technology, Roorkee, Uttarakhand, India
[#]Email: neha88728@gmail.com

Phthalic acid esters (PAES) are routinely used in the large scale production of flexible and durable plastics. PAEs are widely used in commonly used products, ranging from enteric coatings of medication pills to nutritional supplements[1]. Among the PAEs, di(2-ethylhexyl) phthalate (DEHP) is frequently used as an additive in the manufacturing of polyvinyl chloride. Many studies have demonstrated that PAEs such as DEHP resulted in inhibition of important enzymes in essential amino acid metabolism pathway. It has been reported that administration of large amount of Di butyl phthalate (DBP) resulted in increased ratio of conversion of trp to niacin pathway. Inhibition of important enzymes related to this metabolism leads to the formation of excessive quinolinic acid (QA) leading to adverse health effects. Structural insights related to DEHP and its metabolites interaction with important enzyme in essential amino acid metabolism, decipher their

toxic potential. In the present study, enzyme of amino acid metabolism complexed with its substrate analogue was taken and minimized using Discovery studio 4.1. It was subsequently docked with DEHP and its metabolites using Autodock tools 1.5.6 and binding affinity of phthalate metabolites (4.8- 7.2 Kcal/Mol) obtained was in comparable range with that of substrate analogue (-6.2 Kcal/Mol). Molecular-level insights associated with ligand binding were provided through the use of molecular dynamics (MD) simulations (10ns) of different phthalate complexes. Results show very less changes in patterns of backbone RMSD, fluctuation, and compactness, in phthalate complexes as compared to DPA and ACMSD complex. It is concluded that the inhibition of enzyme by DEHP and its metabolites caused increase in the conversion ratio resulting in disturbed tryptophan-niacin metabolism.

P- 50

Prediction of change in aggregation rate upon mutation using support vector machine

Puneet Rawat[1,#], Sandeep Kumar[2], M. Michael Gromiha[1,*]

[1] Department of Biotechnology, Indian Institute of Technology Madras, Chennai, India
[2] Biotherapeutics Pharmaceutical Sciences, Pfizer Inc, Chesterfield, Missouri, USA
E-mail: puneet021192@gmail.com

Multiple cellular processes such as lysosomal degradation and chaperones closely regulate protein aggregation and maintain proteostasis in higher organisms. However, mutations, environmental stresses, and aging can impair their process, leading to accumulation of protein aggregates and amyloids within and outside the cells. *In-vitro* aggregation experiments have shown that even a single amino acid substitution in a protein sequence can drastically increase/decrease its rate of aggregation. In this work, we have collected such examples from literature and classified the mutations into two categories based on changes in protein aggregation rates and identified the sequence-based features that underpin these changes. We have developed a Support Vector Machine (SVM) model using a dataset of 222 mutations in 25 proteins and obtained a prediction accuracy of 69.37%. The classification of mutants into a helix, strand, and coil based on their location improved the prediction accuracy up to 84.5%, 82.8%, and 82%, respectively using leave one out cross-validation. Further, the features selected are distinct in different secondary structures. Specifically, stability and flexibility changes in helical mutations; β-strand propensity, polarity and charge in mutations in β-strand; and aggregation propensity, helical tendency and tendency to form secondary structure in coil structural classes. This suggests that amyloidogenic proteins might aggregate differently depending on secondary structures present at the site of the mutation. This analysis can help us to develop either mutant with less aggregation rate for proteins related to neurodegenerative diseases or mutants with higher aggregation rate for their commercial application as nano-material, drug delivery vehicle etc.

P- 51

Novel splice variation based approach towards mining putative riboSNitches in MAPT pre-mRNA

Ramya G[1,#*], Abhijit Mitra

Center for Computational Natural Sciences and Bioinformatics International Institute of Information Technology, Hyderabad Gachibowli, Hyderabad 500032, Telangana, India
[*]ramya.gurrapu@research.iiit.ac.in

RiboSNitches are single nucleotide polymorphisms (SNPs) which cause significant structural disruption in the pre-mRNA secondary structure ensemble leading to disease phenotype (Halvorsen *et al*, 2010). Till date they have been identified in the 5′ and 3′ untranslated regions of genes associated with the pathogenic conditions.

It has been reported (Warf *et al*, 2010) that in different types of alternative splicing events, variants are seldom related to pre-mRNA secondary structure. We worked on the hypothesis that the SNPs which lead to structural changes in pre-mRNAs may result in differential expression of splice variants. Our study focuses on lethal mutations and SNPs related to alternative splicing

and structural methods based on riboSNitch mining tools. Accordingly, we examined the alternative splicing mechanism of exon 10 in MAPT/ *tau* gene, which is crucial for maintaining the natural balance between three and four repeat protein isoforms. It has been reported earlier (Ghetti *et al*, 2015) that disruption of this natural balance may result in differential expression of the MAPT isoforms, which in turn may lead to neurodegeneration. By pursuing this method it is possible to detect putative riboSNitches which may have been missed in the database of pathogenic SNPs identified by empirical correlation studies such as GWAS. Here we report three putative riboSNitches, which were not found in available pathogenic SNP databases. It is possible that these riboSNitches may be responsible for pathogenicity, in conjunction with other regulating factors.

This method of analyzing SNPs which affects the alternative splicing on the basis of structural changes, without the apriori knowledge regarding their pathogenicity, opens up a new approach towards riboSNitch prediction. Our analysis prepares the ground for designing further investigations for the experimental validation under *in vivo* conditions.

References

[1] Ghetti, Bernardino, *et al.* "Invited review: frontotemporal dementia caused by microtubule associated protein tau gene (MAPT) mutations: a chameleon for neuropathology and neuroimaging." *Neuropathology and applied neurobiology* 41.1 (2015): 24-46.

[2] Halvorsen, Matthew, *et al.* "Disease-associated mutations that alter the RNA structural ensemble." *PLoS genetics* 6.8 (2010): e1001074.

[3] Warf, M. Bryan, and J. Andrew Berglund. "Role of RNA structure in regulating pre-mRNA splicing." *Trends in biochemical sciences* 35.3 (2010): 169-178.

P- 52

Transthyretin – A Novel Therapeutic Target For Breast Cancer

Saurabh Sharma [1,2#], Vandana Mishra[1], A. S. Ethayathullah[2], Prakarsh Yadav[2], B. D. Banerjee[3], Radhey Shyam Sharma[1]

[1]Department of Environmental Studies, Delhi University, Delhi, India
[1]Department of Biophysics, All India Institute of Medical Sciences, New Delhi, India
[1]Department of Biochemistry, University College of Medical Sciences, University of Delhi, Delhi 110 095, India.
E-mail: saurabh1987sharma@gmail.com

Breast cancer (BC) is the second most common cancer diagnosed worldwide and more than 1.3 million women have been diagnosed with BC each year. Lack of effective therapeutic target and poor quality of patients' life after radical mastectomy and chemotherapy are the two major challenges for cancer biologist. Better diagnosis and adjuvant therapy are keys to minimize the severe negative effects of BC on quality of life among the affected women. A serological marker would have great diagnostic value in determining disease progression and the monitoring of therapeutic efficacy because of its greater accessibility over other body fluids based markers. Thus, the aim of this study was to discover more potent serological therapeutic target to arrest the inflammation in initial state and ameliorate the disease condition.

To identify potential serological target, two-dimensional gel electrophoresis (2-DE) was used to analyze differential protein expression in BC patients as compared to the healthy controls. Based on 2-DE and MALDI-TOF/MS-MS analysis, 17 differentially expressed protein spots are recognized. The differential expression of transthyretin (TTR) protein was further validated using Western blot and confirmed by ELISA analyses. Though, TTR primarily involved in the carrier of thyroxine and retinol but it has been suggested as a biomarker for number of diseases such as alzheimer's disease, ovarian cancer, rheumatoid arthritis, malnutrition etc. Further, to identify a natural inhibitor having high affinity with TTR, molecular docking study was performed using known anti-cancer compounds. The docked states were computationally validated using Molecular dynamics (MD) using GROMACS.

A systematic investigation on identification of new serological therapeutic targets would help to arrest disease initiation and progression.

P- 53

Structural and mechanistic insights into interactions of natural anticancer pigments with different cancer macromolecules *in silico*

Vikas Sharma[1,2*#], Prabodh Chander Sharma[1], Vipin Kumar[1,3]

[1]Institute of Pharmaceutical Sciences, Kurukshetra University, Kurukshetra-136119, Haryana, India, [2]University Institute of Pharmaceutical Sciences, Chandigarh University, Gharuan, Mohali, Punjab-140413, [3]Department of Pharmacy, School of Chemical Sciences and Pharmacy, Central University of Rajasthan, Ajmer-305801, Rajasthan, India.

Apart from providing color to plants and animals, pigments also have medicinal properties like antimicrobial, anticancer *etc*. Tetrapyrrole is a class of natural pigment composed of four pyrrole nucleus[1]. Phycocyanobillin, phycourobillin, phycoerythrobillin are natural pigments possessing tetrapyrrole nucleus having anticancer properties but their well-defined mechanism is still to be reported. In view of these, docking investigations were carried out against different cancer macromolecules *viz*. CDK-6, CDK-2, DNA-Quadruplex, BCl-2, VEGFR-2 and IGF-1R kinase, and Type IIA DNA topoisomerase, where alteration of expression for each macromolecule corresponds to a different anticancer mechanism. Different modules of Schrodinger drug design software, *i.e.* Maestro, Ligprep, Protein preparation wizard, Glide were utilized in this study. The obtained results were assessed on parameters like Glide score, hydrogen bonding, docking poses, and different energies like coulomb energy, vander wall energy *etc*. Tetrapyrroles displayed Glide score in the range of -0.02805 to -6.55346. Phycoerythrobillin, phycocyanonbillin and phycourobillin displayed glide scores of -6.23 and -6.55 towards CDK-2 macromolecule respectively whereas phycoerythrobillin exhibited glide score of -6.026 against CDK-6 macromolecule. Phycocyanonbillin showed H-bond interactions with Lys 129 (2.439), Asp 145 (1.688), THR 14 (2.334) residues while interacting residues for phycourobillin were: Asp 86 (1.809), Ash 132 (2.466), Lys 33 (1.668), Leu 83 (1.887). On the other hand, phycoerythrobillin showed H-bond interactions with Gln 149 (2.110), Asn 150 (2.007) residues of CDK-6 macromolecule. The present study discussed the different interactions between the tetrapyrroles and cancer macromolecules. The study concluded that CDK-2 could be the best possible target for phycocyanobillin and phycourobillin whereas CDK-6 could be the preferential target for phycoerythrobillin.

P- 54

Hydration Water Dynamics Around Protein Surface: An Influence of External Potential

Shivangi Sharma[#] and Parbati Biswas[*]

Department of Chemistry, University of Delhi, Delhi -110007. E-mail: pbiswas@chemistry.du.ac.in

A stochastic noise-driven dynamic model is proposed to study the diffusion of water molecules around the protein surface, either under the effect of the thermal fluctuations or under the simultaneous effect of the thermal fluctuations and the effective non-linear potential. Thermal fluctuations arises due to the collisions of the water molecules with the surrounding environment, whereas the effective non-linear harmonic potential emanates from the internal conformational fluctuations of the protein residues, that is transmitted to the surrounding water molecules through dynamic coupling. The dynamics of such a system may be described in the framework of generalized Langevin equation[1], where the thermal fluctuations are assumed to be algebraically correlated[2] in time, which governs the non-Markovian behavior of the system. Results of MSD calculations for the hydration water dynamics for both the cases show a transition from ballistic to subdiffusive behavior, that ranges from short to long time regime for the dynamics considered under thermal fluctuations whereas, it ranges from short to intermediate time regime in presence of potential. Moreover, the subdiffusive behavior obtained in intermediate time regime is not completely demarcated for the different

complexities of the environment, which is otherwise attained in the absence of potential. VACF shows a much faster decay in dynamics of hydration water in the presence of harmonic potential as compared to a potential free case. As the motion of water molecules show a pronounced damping due to the presence of potential and the damping further increases with increase in frequency of harmonic potential. Analytical expressions for first passage time distribution and survival probability are also derived for different boundary conditions, to analyze the hydration water dynamics. The results depict a unimodal distribution of first passage time unlike Brownian motion[3] and survival probability follows a stretched exponential decay for the dynamics of hydration water under thermally correlated noise.

References

[1] Kubo. R.; *Rep. Prog. Phys.*, **29**, 255 (1966).

[2] Rice, S.A.; Dinner, A. R.; *Adv. Chem. Phys.*, **150**, 187 (2012).

[3] Chandrasekhar, S.; *Rev. Mod. Phys.*, **15**, 1(1943).

P- 55

Identification of Heat related SSRs in wheat (*Triticum aestivum*)

Shivangi Varshney[#], Jyotika Bhati, Krishna Kumar Chaturvedi, Anil Rai, Ranjeet R. Kumar, Suneha Goswami, Viswanathan Chinnusamy, Dwijesh Chandra Mishra*

E-mail: shivangivarshney@live.com, singh.jyotika@gmail.com, kk.chaturvedi@icar.gov.in, anilrai64@gmail.com, ranjeetranjaniari@gmail.com, suneha08@gmail.com, viswanathan@iari.res.in, dwijesh.mishra@icar.gov.in

Wheat (*Triticum aestivum) is a very important cereal crop in the word. Its* production is severely limited due to various environmental stresses. Particularly heat stress adversely affects the production of wheat during reproductive development and grain-filling stages. Nowadays, a variety of molecular markers like simple sequence repeats (SSRs) and single nucleotide repeats (SNPs) have become the markers of choice to understand the molecular mechanism in various plants. Thus the objective of the study was to identify the markers SSRs and SNPs in wheat under heat stress. SSRs and SNPs markers were identified in wheat crop under heat stress transcriptomic data. Two contrast i. e. sensitive and resistant varieties of wheat were selected for marker detection using computational approach. Total number of 2129 and 2196 SSRs markers were identified in normal sowing and delayed sowing plants. Five major classes of SSRs that is, mono-, di-, tri-, tetra- and pentanucleotide repeats were targeted for identification of SSRs. These SSR markers have extensive range of applications, like genome mapping, phenotype mapping, marker assisted selection, a range of molecular ecology and diversity studies.

P- 56

In silico comparison of cyclins and cyclin dependent kinases of *Leishmania* and *Plasmodium* and prediction of their subcellular localisation, function and structure

Upasana Das, Trisha Alu, Devarati Sarkar[#], Paulami Dutta, Santanu Roy*

Post Graduate Department of Microbiology, Acharya Prafulla Chandra College, Sodepur Road, New Barrackpore, Kolkata – 700 131
[#]Department of Biotechnology, Heritage Institute of Technology, Chowbaga Road, Anandapur, Kolkata – 700 107
**To whom all correspondences should be addressed*
E-mail: santanuroy@outlook.in.

Parasitic diseases remain among the major causes of human misery and death in the world today and, as such, are important obstacles to the development of the economically less favoured nations. A vast number of the world population, of which mostly are destitute, are ailed by these diseases. Some of the most dreaded ailments like leishmaniases and malaria are caused by the parasitic protozoa of the genera *Leishmania* and *Plasmodium* respectively. The prevalent drugs that are used to treat these diseases has, for the past several years, become less effective due to the emergence of resistant parasite strains leading to a pressing exigency for novel therapeutic approaches against these organisms.

Parasitic protozoa are among some of the most divergent eukaryotes. However, several key biological pathways have been conserved in these

organisms. As for example, though the cell cycle control machineries of these unicellular entities have many unique features, they include certain key regulatory molecules, such as the cyclins and the cyclin dependent kinases (CDKs), albeit with appreciable distinctive characteristics when compared with their mammalian hosts. This suggests that they can be potential targets for parasite specific innovative drugs.

Herein, we have tried to find out the similarities and the differences between the key players of the cell cycles of these parasites and their mammalian host (human). In addition we have also tried to predict their subcellular localisations and functions *in silico*. We have also attempted to generate modelled structures of these parasite cell cycle proteins.

P- 57

Evaluation of human Cathelicidin LL-37 as a novel inhibitor of Panton-valentine Leucocidin (PVL) of MRSA: An *in silico* approach

Himanshu G. Toor[1#], Jenabhai B. Chauhan[1*]

[1]P. G. Department of Genetics, Ashok & Rita Patel Institute of Integrated Study & Research in Biotechnology & Allied Sciences (ARIBAS), New Vallabh Vidyanagar, Anand-388121

[#]Research Scholar; E-mail: himanshutoor013@gmail.com

**Associate professor and Head*

E-mail: jenabhaichauhan@aribas.edu.in

Methicillin-Resistant Staphylococcus aureus (MRSA) has attained paramount importance as an opportunistic pathogen, responsible for the incidence of diverse nosocomial infections in healthcare facilities worldwide, attributed to its plethora of virulence factors and presence of distinctive drug-resistance mechanisms rendering it non-susceptible against conventional antibiotics, instigating the search for novel therapeutic agents possessing broad-spectrum antimicrobial potential with minimal systemic toxicity to the host upon administration. Anti-microbial Peptides (AMPs) have emerged as the frontrunners fulfilling this criterion. PVL (Panton-Valentine Leucocidin) toxin of MRSA is a bicomponent, synergohymenotropic cytolysin whose structure comprises of two non-associated, secretory water-soluble monomers LukS and LukF. The present study focuses on an *in silico* approach to elucidate the inhibitory potential of the human antimicrobial peptide LL-37 against LukS PV and LukF PV subunits of PVL toxin. LL-37 (ligand) was docked with LukS PV and LukF PV subunits (receptor molecules) using the PatchDock server. The first thousand PatchDock results were refined using FireDock server. FireDock provided a set of 10 best structures, arranged in the increasing order of binding energy. Of the two ligand-receptor combinations evaluated, the best combination was that of the human LL-37 peptide (PDB ID: 2K60) with LukS PV (PDB ID: 1T5R) which had a minimum global energy of -61.82. The binding interactions between them were analysed using PyMol molecular viewer. Six bonds were observed between the human Cathelicidin LL-37 (PDB ID: 2K60) and the LukS PV component of the PVL toxin (PDB ID: 1T5R). The molecular dynamics of these interactions were analysed using PyMol which indicated that LL-37 depicted a strong binding potential with the LukS PV. Thus, it could be hypothesized that LL-37 could act as a potential inhibitory agent against PVL toxin and prevent the oligomerization of the LukS PV and LukF PV subunits, thereby protecting the host from its deleterious. However, further validation by experimental research is needed in order to develop this peptide as a novel therapeutic agent.

References

[1] S. Venugopal and R. Mohan, *In Silico* docking studies of *Staphylococcus aureus* virulent proteins with antimicrobial peptides. International Journal of Pharmaceutical Research and Development. 2012; 3(12): 79-86.

[2] D. Schneidman-Duhovny, Y. Inbar, R. Nussinov and H. J. Wolfson, PatchDock and SymmDock: Servers for rigid and symmetric docking. Journal of Nucleic Acids Research. 2005; 33: W363-W367. doi:10.1093/nar/gki481.

[3] E. Mashiach, D. Schneidman-Duhovny, N. Andrusier, R. Nussinov and H. J. Wolfson, FireDock: a web server for fast interaction refinement in molecular docking. Journal of Nucleic Acids Research. 2008; 36: W229-W232. doi:10.1093/nar/gkn186.

P- 58

Comprehensive assessment of miRNA target prediction tools for human and drosophila melanogaster

Muniba Faiza[1], Khushnuma Tanveer[2#], Saman Fatihi[2], Khalid Raza[2*]

[1]School of Food Science and Engineering, South China University of Technology, Guangzhou, China 410640
[2]Department of Computer Science, Jamia Millia Islamia, New Delhi, India 110025
 # Presenting Author khushnuma.tanveer@gmail.com
**Corresponding Author kraza@jmi.ac.in*

MicroRNAs (miRNAs) are small non-coding RNAs that controls gene expression at the post-transcriptional level through complementary base pairing with the target mRNA, leading to mRNA degradation and blocking translation process. Any dysfunctions of these small regulatory molecules have been linked with the development and progression of several diseases. Therefore, it is necessary to reliably predict potential miRNA targets. A large number of computational prediction tools have been developed which provide a faster way to find putative miRNA targets, but at the same time their results are often inconsistent. Hence, finding a reliable functional miRNA target is still a challenging task. Also, each tool is equipped with different algorithms, and it is difficult for the biologists to know which tool is the best choice for their study. This work presents performance assessment of eleven frequently used miRNA target prediction tools using experimentally validated high confident miRNAs and their targets taken from miRBase for two organisms Human and Drosophila Melanogaster. In human dataset, miRMap showed the highest accurate results amongst the other predictors, followed by the TargetScan; whereas in the D. melanogaster dataset, microRNA tool showed the best performance followed by the CoMiR in the comparison of other tools.

P- 59

Computational analysis of protein-carbohydrate complexes to reveal the interplay between binding and stabilizing residues

N. R. Siva Shanmugam[1#], J. Fermin Angelo Selvin[2], K. Veluraja[3] and M. Michael Gromiha[1*]

[1]Department of Biotechnology, Bhupat and Jyoti Mehta School of Bioscience, Indian Institute of Technology Madras, Chennai 600036, Tamilnadu, India
[2]Department of Physics, Manonmanium Sundaranar University, Tirunelveli, Tamilnadu, India
[3]Vellore Institute of Technology, Vellore, Tamilnadu, India
E-mail id: nrsivashanmugam@gmail.com

Protein-carbohydrate interactions are important for cellular processes. The analysis of residues involved in folding and stability would help to understand the recognition mechanism. In this work, we considered 3,80,216 residues in 1,130 protein-carbohydrate complexes and identified 9,302 binding residues and 26,873 stabilizing residues using distance based criteria, sequence and structural properties. Among them, 5.9% of binding and 2.04% of stabilizing residues are identified as key residues. These key residues are analyzed with the preference in structural classes, preferred amino acid residues, contact between protein and carbohydrate partners, functional classes and structural based parameters. Based on structural classifications, key residues are dominant in all-β, α+β and α/β classes. Residue-wise analysis showed that polar and charged residues have high tendency to serve as key residues. In addition, Lys is preferred in all three gene ontology terms. Atomic level analysis showed that key residues are preferred to form polar-nonpolar contacts followed by the contacts between charged atoms. The salient features of the results will be discussed.

P- 60

Genome Wide Identification and Functional Annotation of Genes in Sesame (*Sesamum indicum* L.)

Supriya Purru[1], Sarika Sahu[1], Rao AR[1], Bhat KV[2*]

[1]ICAR-Indian Agricultural Statistics Research Institute, New Delhi
[2]ICAR-National Bureau of Plant Genetic Resources, New Delhi
**Corresponding author: KV.bhat@icar.gov.in*

Sesame (*Sesamum indicum* L.) is an ancient and important oilseed crop belonging to the family pedaliaceae. Sesame, as a source of high quality oil is valued for its stability, nutritional value and resistance to rancidity and is often referred as the "Queen of oil seeds". The plant has tremendous nutritive value however, intensive studies on genetics and molecular mechanisms responsible for yield and adaptability traits are lacking. Hence, it is essential to explore the sesame plant with state of the art of sequencing technology to improve the quality and yield. In this study, a total of 24,579 genes were predicted from whole genome assembly of Sweta variety of sesame. Blast2Go software was used to assign Gene ontology terms to classify the functions of the predicted genes. The GO terms were classified into 59 functional groups including biological processes (25), cellular component (19), and molecular function (15). In the biological process category metabolic process was the most prevalent (30.4% of sequences), followed by cellular process (28.4%) and response to stimulus (11.2%). In the molecular function category, binding was the most prevalent (46.4%), followed by catalytic activity (39.5%), structural molecule activity (4.95%) and transporter activity (4.95%). In the cellular component category, cell was the most dominant term (3.37%), followed by Organelle (26.45%) and macromolecular complex (6.1%). The predicted genes were found to involve in pathways such as fatty acid biosynthesis, purine metabolism, starch and sucrose metabolism etc. The identified genes in this study will not only facilitate the understanding of genetic basis of fatty acid biosynthesis, but also accelerate genetic improvement through marker-assisted selection in sesame.

P- 61

Label-Free Quantitative Proteomics of Over Expressing Leukemia Inhibitory Factor (LIF) COS-1 Cells Used for Computation Prediction of PTMs Cross Talk

Syed Azmal Ali[#], Ashok Kumar Mohanty, Sudarshan Kumar[*]

Proteomics and Cell Biology Lab, Animal Biotechnology Center, National Dairy Research Institute, Haryana, India
E-mail: kumarsudershan@gmail.com

Leukemia inhibitory factor (LIF) is a polyfunctional cytokine belongs to an IL-6 class of interleukin family. It is ubiquitously present in different organs, but the central mechanism of action exerts by LIF signaling for cell survival and polarity is poorly understood. In the present study, we have applied the high-resolution mass spectrometer (Q-TOF) based label free quantitative proteomics approach for the identification of high abundant proteins under the influence of LIF. Identified high confident 2083 protein were used for the prediction of commonly occurring PTMs, including phosphorylation, acetylation, methylation, sumoylation, S-nitrosylation, and lipid residues binding proteins (palmitoylation, N-myristoylation, farnesylation, and geranylgeranylation). Totally, we computationally identified 23351 sites for acetylation, 25367 sites for phosphorylation, 18453 sites for methylation, 2337 sites for S-nitrosylation, 6113 sites for sumoylation and finally 2337 sites were found to be available for lipid binding. We further identified the 364 common proteins which contain all the PTMs sites used in the analysis (considering at least per PTM one site is available) and these could be utilized by these proteins for their PTM cross talk to exhibit cellular response. Now, through the bioinformatics analysis, we have used the GeneMANIA database to download the PPI network for identified 364 common proteins to predict the PTMs cross talk interaction and the network was created in Cytoscape platform. Integrated network analyzer was used to create node or edge data values and betweenness centrality calculation applied to visualize the central hub proteins in the LIF induced protein network. Through this way, we have identified AKT1,

CDC42, UBA52 and UBC to be central hub proteins in the network containing all the search PTM sites and responsible for LIF/LIFR induced downstream signaling. Further our prediction gets strong strength by the fact that alone AKT kinase is responsible for the phosphorylation of 1494 proteins out of 2083 total identified proteins. Through these computational analyses, we conclude that LIF involves in autocrine-paracrine mediated cell cycle signaling. These identified targets suggest that LIF could be an important prognostic marker for various diseases such as neurodegenerative diseases and cancer.

P-62

Binding of Polychlorinated biphenyl to Hormone receptor: an *in silico* study

Vikram Dalal[#], Neha Singh and Pravindra Kumar*

Departmentof Biotechnology, IIT Roorkee, Roorkee (Uttrakhand), India
E:mail: vikram.dalal.37@gmail.com

Biphenyl is an aromatic hydrocarbon with a molecular formula $(C_6H_5)_2$. It is used in organic synthesis, heat transfer fluids, dye carriers, food preservatives and as a fungistat in packing of citrus fruits. Polychlorinated Biphenyls (PCBs) are manufactured by the direct chlorination of biphenyl. Polychlorinated biphenyls (PCBs) are industrial compounds that have been detected as contaminants in almost every component of the global ecosystem including the air, water, sediments, riu, and wildlife and human adipose tissue, milk, and serum[1]. The PCBs can increase the risk of developing cardiovascular, liver disease and diabetes. Women are at high risk of giving birth to infants of low birthweight, who are at high lifetime risk for several diseases. Mono-hydroxylated polychlorinated biphenyls(OH-PCBs) are found in human biological samples and lack of data on their potential estrogenic activity has been a source of concern. The present study was carried out to investigate whether biphenyl congeners can bind to human hormone receptors and elucidate its mode of interaction. Geometric optimization of the protein was done using Discovery Studio 4.0. Biphenyl congeners have been successfully docked with hormone receptor into the ligand binding site using AutoDock Tools1.5.6. It has been found that the mode of interactions and binding energy of biphenyl congeners with hormone receptor were similar to that of hormone. Further, the stability of receptor-biphenyl congeners complex was validated using molecular dynamics simulation studies in Gromacs 5.1.4[2]. Molecular dynamics (10ns) revealed similar pattern of deviation, fluctuation and compactness in biphenyl congeners complex compared to the hormone receptor containing hormone. While intermolecular hydrogen bonding, ligand deviation and solvent accessible surface area showed smaller changes in biphenyl complexes as compared to hormone receptor complex. Our study focuses on the importance of computational methods related to toxicity of biphenyl congeners binding to hormone receptors.

Reference

[1] Carpenter DO. Polychlorinated biphenyls (PCBs): routes of exposure and effects on human health. Reviews on environmental health. 2006; 21(1): 1-24.

[2] Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX. 2015 Sep 30; 1: 19-25.

P-63

A computational approach to evaluate the deleterious non-synonymous variants and protein phenotype prediction of Human *ARHGEF6* gene

Yashvant M. Khimsuriya[1#], Jenabhai B. Chauhan[1*]

[1] P. G. Department of Genetics, Ashok & Rita Patel Institute of Integrated Study & Research in Biotechnology & Allied Sciences (ARIBAS), New Vallabh Vidyanagar, Anand-388121
[#] Research Scholar, Email: yashvantkhimsuriya@aribas.edu.in
**Associate Professor & Head*
[*]E-mail: jenabhaichauhan@aribas.edu.in

The genetic mutations in *ARHGEF6*, a guanine nucleotide exchange factor that activates the Ras-like family of Rho proteins via GPCR (G-protein-coupled receptor) pathway are reported to cause non-syndromic X-linked intellectual disability and

supraglotis cancer [1]. However, the fundamentals of how these genetic mutations draw out the protein's structure and functional effect remains unknown. Therefore the present study aimed to scrutinize the consequences of ARHGEF6 mutations, using various Bioinformatics methods [2]. At first, the ensemble Genome Browser was used to classify the different classes of mutations of ARHGEF6 gene based on their pathogenic impact. The functional and disease effect of non-synonymous variants were predicted. The protein structures for wild and mutant types built by MUSTER. These structures were examined for protein stability, structural divergence, solvent accessibility and functional interaction deformities. A total of 169 missense and 13 non-sense variants were found in *ARHGEF6* (Ensembl transcript: ENST00000250617.6). Out of these, 47 missense variants were predicted to be damaging and deleterious effect commonly evaluated by SIFT, POLYPHEN2 and PROVEAN scores. Then, 21 of them were predicted to be less stable by I-muatant2.0 and MutPred2. Finally, this study was able to point out the genetic mutations mapped to total 12 exons out of 22 exons of *ARHGEF6 gene*, highly conserved CH , SH3, DBL homology, PH and Rho guanine nucleotide exchange factor domains which alters the protein phenotype as well as the stability of ARHGEF6-GIT2 interactions and 9 more protein interactions. The present study findings may narrow down the number of pathogenic mutations of *ARHGEF6* gene to be study for Rho guanine nucleotide exchange factor associated genetic diseases. This study on *in silico* analysis is a basic approach to prioritize the rational impact of pathogenic mutations.

References

[1] Lower K, Gecz J. "Characterization of ARHGEF6, a guanine nucleotide exchange factor for Rho GTPases and a candidate gene for X-linked mental retardation: mutation screening in Börjeson-Forssman-Lehmann syndrome and MRX27." Am J Med Genet 2001; 100: 43–8.

[2] Hussain MRM, Shaik NA, Yousuf Al-Aama J, Asfour HZ, Khan FS, Masoodi TA, et al. In silico analysis of Single Nucleotide Polymorphisms (SNPs) in human BRAF gene. Gene 2012; 508: 188–96.

P-64

I-TRAQ proteomic and biological network analysis revealed the molecular mechanism of radioprotection offered by radioresistant bacterial metabolite RKIP-006 in c57bl/6 mice

Shravan Kumar Singh, Ashutosh K Gupta, Neha Chhachhia, Darshna Singh, Raj Kumar*

Division of Radioprotective Drug Development and Research (RDDR), Institute of Nuclear Medicine and Allied Sciences (INMAS-DRDO), Brig. S.K. Mazumdar Road, Delhi-110054
**E-mail: rajkumar790@yahoo.com*

Planed (medical interventions) or unplanned (Nuclear accidents) ionizing radiation exposure altered the proteome homeostasis and physiology that manifest to radiation injury including lethal and sub-lethal hematopoietic & gastrointestinal acute radiation syndrome. Few prospective radiation countermeasures are still under developmental stage worldwide and actively investigated in terms of dose, timing of administration and mechanisms of its action.

In the present study, we were investigated the molecular mechanisms ofradioprotection offered by radioresistant bacterial metabolite RKIP006 in c57bl/6 mice using techniques of iTRAQ quantitative proteomics and bioinformatics.

Radioprotective efficacy of RK-IP-006 was evaluated using c57bl/6 mice though oral route administration. Quantitative proteomics analysis based on iTRAQ technology was performed to determine the protein profile changes in the jejunum part of small intestine of c57bl/6 mice treated by gamma radiation (10Gy) and radioresistant bacterial metabolite RKIP-006. The obtained LC-MS/MS data were analyzed and grouped under gene ontology categories cellular localization, molecular function and biological process.

Bacterial metabolite RKIP006 showed significant radioprotective efficacy in c57bl/6 mice. Using iTRAQ technology, a total of 298 differentially expressed proteins (142 up regulated & 147 down regulated) in RK-IP-006 treated plus irradiated c57bl/6 mice were identified and subjected to gene ontology analysis. Gene ontology

analysis showed the differentially expressed proteins were mainly associated with immune response (S100A8; Integrin alpha-IIb/beta-3), cytoskeleton (Vinculin, Alpha-actinin-1), DNA repair & chromatin modulation (ATP-dependent DNA helicase 2 subunit 1, Histone H2A type 2-B) and signaling (14-3-3 protein beta/alpha, platelet glycoprotein 4) biological process. Protein interaction network and KEGG pathway analysis revealed altered metabolic pathways, glycolysis, TCA and antioxidant pathways.

Altogether, it can be concluded that multiple pathways (network of proteins) involved in the radioprotection and bacterial metabolite RKIP006 pretreatment overcome radiation induced injury by improving DNA repair & chromatin, cytoskeleton modulation, immune response and regulation of signal transduction.

P-65

Genome-wide identification, characterisation and validation of drought responsive CDPK gene family in *Arabidopsis*, maize, rice and sorghum

Shikha Mittal[3#], Mallana Gowdra Mallikarjuna[1], Prashant Ankur Jain[2], Atmakuri Ramakrishna Rao[3], Nepolean Thirunavukkarasu[1*]

[1]Division of Genetics, ICAR-Indian Agricultural Research Institute, New Delhi-110012, India; [2]Department of Computational Biology & Bioinformatics, Sam Higginbottom University of Agriculture, Technology and Sciences, Allahabad- 211007, India; [3]Centre for Agricultural Bioinformatics, ICAR-Indian Agricultural Statistics Research Institute, Pusa, Library Avenue, New Delhi- 110 012, India

Calcium dependent protein kinases (CDPKs) plays significant role in regulation of plant growth, development and response to various stresses. Comparative and evolutionary analyses of CDPK gene family among *A. thaliana*, *Z. mays*, *O. sativa* and *S. bicolor* was performed using various *in-silico* tools. Expression analysis of drought responsive CDPK genes was carried out using microarray expression datasets. A total of 32, 72, 78 and 91 CDPK genes were identified in maize, *Arabidopsis*, rice and sorghum, respectively. The phylogenetic tree was divided into 4 groups, in which maize CDPKs were observed only in group III while other species were well distributed in all groups. The ratio of non-synonymous and synonymous substitution between *Z. mays* and *A. thaliana*, *Z. mays* and *O. sativa,* and *Z. mays* and *S. bicolor* were 0.01, 0.07 and 0.08, respectively, which indicated that they had undergone natural selection. Ramachandran plot analysis of 3D protein structures showed approximately all the amino acids in favoured and allowed regions suggesting good prediction accuracy. Out of 278 genes, 80 genes were identified as drought responsive genes using expression data in root, stem and leaf tissues at different time points and duration. Maximum number of drought related expression was noticed in leaf tissue, followed by root and stem. The information generated on CDPK genes could be used in development of drought tolerant genotypes and transgenic using appropriate breeding tools.

P-66

Next-Generation Sequencing (NGS) of Genome – Emerging Technologies and Applications

Pallavi Mishra[1#*], Ranjeet Maurya[2], Himanshu Avashthi[1], Amrender Kumar[1] and A. K. Mishra[1]

[1]Agriculture Knowledge Management Unit, ICAR-IARI, New Delhi, India; [2]Department of Molecular biology and Genetic Engineering, G.B. Pant University of Agriculture and Technology, Pantnagar, Uttarakhand, India
**E-mail: mishrapallavi58@gmail.com*

Next-generation sequencing (NGS) is possibly the most dynamic and rapidly growing area of modern biology. The aim for the technology will be to lower the cost of the equipment and biochemicals involved, increasing simultaneously the reproducibility, reliability and simplicity of the techniques and protocols in operation. During the last few years there is rapid development in sequencing technology, innovative biological applications, and softwares. The advancement of Next-Generation Sequencing (NGS) technology facilitates generation of enormous amount of raw data which needs to be analyzed to gain meaningful results and for knowledge discovery. The transition from the first human genome sequences to the personal genomes and genomic solution for

medicine and agriculture has been made possible only because of the advances in sequencing technologies over the past 13 years. Since the introduction of NGS technology, major transformation in the process of extracting genetic information from biological systems has been revealed limitless insight about the genome. This ability has catalyzed a number of important breakthroughs, advancing scientific fields from human and plant disease research to evolutionary science. As a result new software tools for NGS data analysis are often released the cost of sequencing continues to fall and the scope of sequencing projects expands to larger studies and advances into new areas. Data processing and big data analysis in crop remain key challenges, which is enabled by sophisticated and novel bioinformatics tools. Apart from this not only new software has been developed for a wide range of novel applications and types of data analysis, but new algorithms have also been developed for old problems and to cope with the huge volumes of data generated on new sequencing machines.

P-67

Identification of important structural regions and validation in HopS2 effector protein

Sapna Mayuri Borah#, Dr. Anupam Nath Jha*

Department of Molecular Biology and Biotechnology
Tezpur University, Napaam, Tezpur
*E-mail: sapna@tezu.ernet.in, *anjha@tezu.ernet.in*

Type III secretion systems (T3SS) are molecular machineries in gram negative bacteria that help in translocation of virulent effector proteins into their respective host. They serve as efficient models to understand the mechanisms of host-pathogen interactions. Moreover, these effector proteins may serve as potential targets for controlling the pathogenicity conferred by the respective bacteria. Present literature review reveals much less is known about these effector proteins. Our understanding about these proteins is also limited due to the lack of sequence-structure-function relationship among these effector proteins.

In this study, we have selected one such T3S effector protein HopS2, which is secreted by phytopathogen *Pseudomonas syringae* pv. tomato DC3000. A cognate chaperone shcS2 (specific hop chaperone S2) has also been identified to help in the translocation of HopS2. Here, we elucidate an *in-silico* method to analyse few structural attributes of the protein. We have performed sequence analysis, *ab-initio* structure prediction and validation followed by identification of important structural regions of the protein. While each of the steps aim at searching for significant signatures in HopS2, the findings from available information also led us to a few probable regions that contribute to the effector-chaperone interactions. The selected predicted models have been validated and undergo a total of 800ns molecular dynamics simulation. The stability of the selected regions predicted to be vital has been assessed from the MD trajectory analysis. Different secondary structure determining methods have been used to calculate the occurrence of these local secondary structures during the simulations.

The analysis has provided insight on important functional regions that may be crucial to the effector structure and functioning. The results will aid in experiments to attain and evaluate the structural and functional aspects of this protein family.

References

[1] Hueck, C. J., 1998. Type III protein secretion systems in bacterial pathogens of animals and plants. *Microbiology and molecular biology reviews*. *62* (2), 379-433.

[2] Jayaram, B. *et al.*,2014. Bhageerath-H: A homology/ ab initio hybrid server for predicting tertiary structures of monomeric soluble proteins. *BMC bioinformatics*.*15*.

[3] Portaliou *et al.*, 2016. Type III Secretion: Building and Operating a Remarkable Nanomachine. *Trends in Biochemical Science*. 41(2), 175-89.

P-68

Challenges in Visualization of Biological Network

Rashmi Rameshwari[*], Dr. Shilpa S Chapadgonkar[*] and Dr. T. V. Prasad[**]

[*]Dept. of Biotechnology, Manav Rachna International University, Faridabad,Haryana
[**]Godavari Institute of Engineering and Technology, Rajahmundry, Andhra pradesh
E-mail: [*]rashmi.fet@mriu.edu.in, [*]shilpas.fet@mriu.edu.in, [**]tvprasad2002@yahoo.com

Living systems can be envisioned as a beautiful fabric with weaves of different biological networks such as Genetic, Protein, transcription factors and metabolic interactions. The weave pattern or the network architecture is a precise predictor of healthy or diseased state of the organism. Network based approach gives insight into pathogenesis pathway which leads to drug discovery process. It helps researcher and clinician in grouping together the proteins that interact in functional complexes and pathways. Thus exploring disease networks nodes as potential target for drug discovery. It can also give valuable insight into the possible ways to treat the condition. Presently, our understanding of biological network analysis is expanding with the advent of high through-put technologies. Huge datasets have generated on components of biological networks such as genes, RNAs and proteins and their interactions. Bioinformatics network visualization tools aim to capture the network architecture to aid the interpretation of such datasets. However, biological networks are characterized by complex dynamic behaviour. Visualization tools therefore necessitate a trade-off between the capture of the extent of complexity and comprehensibility. Biological data is often structured in the form of complex interconnected networks such as protein interaction and metabolic networks.

Visualization tools based on different algorithms are available freely for the benefit of scientific community. For example tools for 2D and 3D network visualization tools are freely available, including Cytoscape (*http://www.cytoscape.org/*), APID (*http://bioinfow.dep.usal.es/apid/*) and Networkanalyst (*http://www.networkanalyst.ca*). These tools support multiplatform, inter-operability, visualization quality, visualization features and data coverage. User can upload single or multiple gene expression dataset to perform comprehensive gene annotation. Main challenges to biological network visualization lies in the sheer size, complexity and dynamic nature of networks and overall algorithm responsible for visualization. Visualization tools are based on certain graph theory. At present for visualization, interaction between molecule is shown by directed and undirected graph. To visualize time course behaviour of gene there is need to combine these two graph theory. which can interpret biological phenomenon in better way. Moreover integrating the ever increasing pool of knowledge on components and to filter data to avoid noise are daunting tasks faced by systems biologist. To overcome this a holistic approach is required.

P-69

Development of Information System on Livestock Epigenetics

Sayanti Guha Majumdar[1#] and A. R. Rao[2*]

[1] Indian Agricultural Statistics Research Institute;
[2]Indian Agricultural Statistics Research Institute
E-mail: rao.cshl.work@gmail.com, sayanti23gm@gmail.com

To meet the demand for livestock products, it is essential to understand molecular mechanisms of livestock species. Epigenetics is an emerging field which deals with the study of mitotic and meiotic (or both) heritable changes in gene function that cannot be explained by changes in DNA sequence. In general, three epigenetic mechanisms are available in nature, i.e. (i) DNA methylation (ii) Histone modification and (iii) RNA interference (RNAi). However, the information related to the said epigenetic mechanisms in livestock species is not available at one place and analysis is required for improvement in production traits and controlling diseases in livestock. Hence, the aim of the paper is to (i) parse and analyze data of different epigenetic mechanisms in livestock species, (ii) populate a database on epigenetic mechanism and make it available through web interface. Initially, the gene and protein sequence information related to epigenetic mechanisms of livestock species has

been retrieved from NCBI, UCSC and CABin. Besides, the micro RNA information of cattle and sheep has been retrieved from miRBase. Subsequently, *Sequence Manipulation Suit: CpG Islands* tool is used to obtain the probable methylation sites present in the 1K upstream regions of genes in cattle. In addition, the three dimensional structures of histone proteins of cow, sheep, goat, camel have been predicted, validated, refined and stabilized. Also, the probable genomic regions of histone proteins in buffalo were predicted. Besides, the secondary and tertiary structures of microRNAs of cattle and sheep were predicted. In addition, the miRNA information in buffalo species has been predicted by using miRNAs of cattle. A web-based Information System has been developed with MySQL database as bottom layer, PHP as server side application-middle layer and HTML, CSS and JavaScript at top layer. The developed information system can be accessed at *http://bioinformatics.iasri.res.in/edil.*

References

[1] Bjornsson, H.T., Fallin, M.D. and Feinberg, A.P. (2004). An integrated epigenetic and genetic approach to common human disease. Trends Genet, 20**:** 350–8.

[2] Ke, X., Cortina-Borja, M., Silva, B.C., Lowe, R., Rakyan, V. and Balding, D. (2013). Integrated analysis of genome-wide genetic and epigenetic association data for identification of disease mechanisms. *Epigenetics,* 8: 1236-1244.

[3] Rao, A. R., Dash, M., Sahu, T.K., Wahi, S.D., Behera, B.K., Sharma, A.P. and Bhatia, V.K. (2014). Statistical and bio-computational applications in animal sciences. *Indian Journal of Animal Sciences,* 84 (5): 475–489.

P-70

Drug Repurposing: A Boon for Treatment of Several Types of Cancer

 # Deepshikha Ghosh and Koel Mukherjee*

Bioinformatics Lab, Department of Bio-Engineering, Birla Institute of Technology, Mesra, Ranchi, Jharkhand-835215

Drug Repurposing is based on the principle of "polypharmacology" which literally means one drug and multiple targets. This strategy has evolved from the off-target effects of the existing drugs. *Drug Repurposing is* also commonly known as *Drug Repositioning*, which refers to new manifestations of old drugs. Cervical cancer is the fourth common cancer in women worldwide and ranks no. 2 in the Indian population. *Human Papilloma Virus (HPV)* is the main cause of cervical cancer which encodes the oncoprotein E6. The traditional drug discovery in contrast to drug recycling consumes billions of expenses and time to make a drug consumable. Cancer is a complex disease in which cells in a specific tissue are no longer fully responsive to the signals within the tissue that regulate cellular differentiation, survival, proliferation and death. The purpose of this study is to repurpose native ovarian drugs against some new targets for cervical cancer treatment using the available tools and software's. The 3D models of protein HPV-16 E6 were developed and then drug molecules against HPV Proteins from drug databases were identified. Each of the screened drugs were docked to the active sites of the target protein using the AutoDock tool. Based on the docking scores of different protein-drug complexes ,the drug with the best *Dock Score* was noted as *Rucaparib*. All the protein-drug complexes were optimized and simulated using GROMACS. When E6 binds to p^{53}, p^{53} is inactivated and cannot function in cell cycle control. Rucaparib will be established as one of the repurposed drugs for Cervical Cancer treatment thereby reducing the cost of designing new drugs.

P-71

Amaranthus Microsatellite Database (AMD): Genome level information on frequency and distribution of SSR

Sarika Sahu[1], Rekha Dixit[2], Indra Singh[1], Himanshu Kumar[1], Jaya Pandey[1] and A. R. Rao[1]

[1]ICAR-Indian Agricultural Statistical Research Institute, Pusa, New Delhi
[2]Subharti University Meerut

Whole genome based molecular markers have potential to accelerate crop improvement. Among these, simple sequence repeats, (SSRs) markers are the most widely used type of molecular markers in crops. Whole genome of Amaranthus has been sequenced using the next generation sequencing platform and available in public domain. This genomic information is a valuable tool for developing molecular markers for diversity analysis

and phylogenetic studies. SSRs are important class of genetic markers as they are abundant, hypervariable, multiallelic and evenly distributed throughout the genome. These markers are ideal resource for marker-assisted breeding programs, germplasm analysis, gene tagging and molecular mapping. In the present study, characterization of SSR markers was carried out in the assembled genome of amaranths species (*A. hypochondriacus*). The frequency of SSR markers with respect to the motif length, type and repeat number was determined. Genome wide localization of identified SSR markers was further carried out. A total of 49964 SSRs were identified in the whole genome. The most abundant repeats were tri nucleotide repeats. Frequency of repeats decreased with the increase in the motif length. Both reverse and forward primers have been identified for each SSR. An online information system on microsatellites was developed and made available in public domain. Genomic distribution of microsatellites will provide a useful tool to be efficiently implemented in amaranths breeding programmes and cross species utilization. Ease and economy of using SSR marker system would benefit Amaranthus improvement through marker assisted selection, gene tagging studies and DNA fingerprinting of varieties. More genetic markers are needed for improvement of Amaranths through marker assisted breeding and saturate the existing genetic map.

P-72

Molecular dynamic simulation and In vitro activity of UDP-N-Acetylglucosamine 1-Carboxyvinyltransferase inhibitors against *Mycobacterium tuberculosis* H37Rv strain

***Mustafa Alhaji Isa, Rita Singh Majumdar and Balwant Kishan Malik**

Department of Biotechnology, School of Engineering and Technology, Sharda University, Greater Noida, India-201306

Tuberculosis (TB) is considered as one of the most devastating global public health threats of the 21st century. It is an infectious disease that is responsible for second cause of death, after human immunodefiency virus (HIV). Multi-drug resistant tuberculosis (MDR-TB) poses grave challenge because of prolonged, limited and expensive treatment options with 10 to 30 per cent of cases resulting in failure of treatment and death. The aimed of this study was to identify novel inhibitors of UDP-N-Acetylglucosamine 1-Carboxyvinyltransferase using both natural and synthetics ligands libraries. To determine the 3D structures of the enzyme and their structural features, homology modeling was used using Modeller9.17. The modelled structure was further checked for high reliability by verify score and Ramachandran plot. Thegenerated model was used for molecular docking simulation studies for predicting the best inhibitors, the selected inhibitors were subjected to absorption, distribution, metabolism, excretion and toxicology (ADME/Tox) prediction, molecular dynamic simulation and In vitro studies. Twenty (20) compounds possessed high activities with minimum free binding energy ranges -10.73 to -8.76kcal/mol. Among the 20 compounds eight displayed the best conformation fitting stability in the binding sites of UDP-N-Acetylglucosamine 1-Carboxyvinyltransferase analysis by molecular dynamics simulation for 5ns using Amber10. The compounds were subjected to bioactivity validation using the resazurin microtiter assay. Among the compound 3 showed high inhibitory activity up to 85% at 10 µg/mL concentration against the growth of the *Mtb* H37Rv strain. Therefore, the identified ligands would serve better lead compounds for future drug design perspective of both multidrug resistance and extensive drug resistance of *Mtb* H37Rv strain.

P-73

2D-QSAR and soft docking: an integrative approach for screening of phytochemicals

Nikita Bora[#], Dr Anupam Nath Jha[*]

Department of Molecular Biology and Biotechnology Tezpur University, Napaam, Tezpur
E-mail: boranikita24@gmail.com, anjha@tezu.ernet.in

Integration of computational strategies and biological knowledge has paved the way for screening out therapeutics against various diseases. Here we have applied different computational techniques to virtually screen out molecules against the agonist conformation of the Adenosine A2A

receptor involved in the modulation of inflammation and insulin resistance. A dataset of 143 phytochemicals has been created belonging to different categories like Flavonoids, Alkaloids, Terpenes, Sulfonylurea.

Initially 118 molecules showing druglike properties have been screened. 2D QSAR study involving a Multiple Linear Regression analysis was performed to estimate the correlation between the physicochemical properties and the experimental values of LD_{50}. Less toxic molecules predicted through QSAR have been selected. Six bioactive compounds satisfying ADMET properties were considered for further analysis.

A multi-template homology modeling has been applied to model the missing residues of the target protein. Minimized conformation of both the modelled receptor and the ligands has been considered to identify the bound conformations of the ligand to the receptor. The binding affinity has been estimated through different docking approaches: i) rigid docking ii) with an explicit hydration to ligands iii) flexibility at active site. Through water mediated docking simulations, we have analysed the binding behavior in presence of water molecules. The possible conformational space of the active site has been explored by keeping the binding site residues as flexible. The stability of these docked complex structures has been evaluated through quantifying the non-covalent interactions. It has been observed that the reported active site residues are involved in the intermolecular interactions.

This integrative approach has revealed that 4 selected compounds might be efficient inhibitors for adenosine A2A receptor. The molecules screened through the above method might be considered for *in-vitro* experiments.

References

[1] Carpenter *et al.* 2016. Structure of the adenosine A2A receptor bound to an engineered G protein. *Nature*, 536 (7614): 104-7.

[2] Forli *et al*, 2016.Computational protein–ligand docking and virtual drug screening with the AutoDock suite. *Nature Protocols* 11, 905–919.

P-74

Association of Psoriasis and Metabolic Syndrome: An Integrative R based Meta-Analysis

Saumya Choudhary[1, 2#], Dibyabhabha Pradhan[2], Sheeba Khan[3], Shivani Rustagi[4], Arun Kumar Jain[2*], George Thomas[1*]

[1] Dept. of Molecular & Cellular Engineering, Sam Higginbottom University of Agriculture, Technology & Sciences, Allahabad, India 211007; [2] Biomedical Informatics Centre, National Institute of Pathology, ICMR, New Delhi, India – 110029; [3] Warner College of Dairy Technology , Sam Higginbottom University of Agriculture, Technology & Sciences, Allahabad, India – 11007; [4] Amity Institute of Food Technology, Amity University, Noida, India – 201313
*E-mail: #saumyachoudhary.biotech@gmail.com, *drakjain@gmail.com, *georgethomas@shiats.edu.in*

Psoriasis is a chronic inflammatory disorder. With inconsistent results, previous reports have implicated the prevalence of metabolic syndrome in psoriasis. We sought to determine the prevalence with an overview and statistical summary of the previous literature with elucidating sub-group analysis. This was a meta-analysis of case-control studies of psoriasis in Indian scenario using random effect model using open R platform. A systematic search of observational studies of psoriasis as study variable and metabolic syndrome and associated risk factor as outcome, published before March 2017 was conducted. Of 15 references in the original search, 11 relevant articles were identified. We included 1115 psoriasis patients and 1168 control. Meta-analysis of extracted studies estimated ratio of true heterogeneity to total observed variation, (I^2 =65%) and between study variance (σ^2) was 0.2555. Using random effects analysis to account for study heterogeneity, the pooled OR for metabolic syndrome was 2.45 (95% CI, 1.71-3.52 and p-value < 0.01). Sub-group analysis for co-related associated factor, hypertension (OR 1.85; 95% CI, 1.33 - 2.57), raised triglyceride (OR 1.95; 95%CI, 1.26 - 3.02), fasting plasma glucose (OR 1.85; 95%CI 1.33 - 2.57) showed strong association whereas waist circumference (OR 1.29; 95% CI 0.97 - 1.73) and HDL cholesterol (OR 1.32; 95%CI 0.73 - 2.39) elucidated weak association in psoriasis patients. In aggregate, our study advocates strong correlation of metabolic syndrome in patients of psoriasis. No single study, whether meta-analytic or not, will provide the

definitive understanding of responses to treatment, diagnostic tests, or risk factors influencing disease. Despite this limitation, meta-analytic approaches have demonstrable benefits in addressing the limitations of study size, can include diverse populations, provide the opportunity to evaluate new hypotheses, and are more valuable than any single study contributing to the analysis.

P-75

Optimization of Rate of Reaction Using Nature Inspired Optimization Algorithms

Akash Shrivastava*, Jagmeet Kaur, Abhishek Kapoor, Shatrughan Modi and Prashant Singh Rana#

Computer Science and Engineering Department, Thapar University, Patiala, India
E-mail: akashshrivastava0406@gmail.com, jagmeetkaurtiet@gmail.com, kapoorabhishek43@gmail.com, shatrughan.modi@thapar.edu and psrana@gmail.com

Chemical reaction involves transformation of reactant into product. Rate of chemical reaction is the measure of how fast these changes are taking place. Some reactions occur very rapidly, others very slowly. For example, ionic reactions are very fast, while those taking place in water treatment plant may last up to few days. Rate of reaction can also be defined as the speed at which a reaction happens. If a chemical reaction has high rate, shows that molecules combine at a higher rate than the reaction has slow rate. The rate of chemical reaction can also depends on different types of molecule that are combining. If there are low concentrations of an essential element or compound, the reaction will be slower.

Earlier Arrhenius equation was used to calculate the rate of reaction with some parameters. But to know the exact reaction mechanism, actual reaction needed to be executed, which was not feasible and time consuming. There are some simulators available for chemical kinetics which can compute the reaction rate but not in optimal way.

This paper focuses on the optimization of rate of chemical reaction using four different Nature Inspired Optimization Algorithms i.e. Random, Genetic Algorithm, Differential Evolution and Particle Swarm Optimization to maximize the rate of chemical reaction and to get the best possible reaction rate along with the input tuning parameters. Test results shows that different algorithms perform significantly better for different reactions and have different convergence rate. In this paper, a simulator is used which an object-oriented software tool named Cantera to calculate the rate of reaction which uses modified Arrhenius equation.

With the help of simulators, we need not to perform the actual reaction to know the reaction mechanism. These simulators can give the reaction rate but not in optimal way.

Optimization algorithms are applied along with the simulation in order to get the best possible reaction rate. We can also determine what should be the appropriate set of inputs for some desired reaction rate with the help of optimization algorithms. After applying all four algorithms over simulated data of reactions from cantera, we found that Genetic algorithm performs much better than all other algorithms in reaction 1, 2 and 4. In reaction 3, DE outperforms all the algorithms but GA has total best score.

P-76

Peptidomics: A Way Forward in Development of Functional Foods.

Sheeba Khan[1#], Shivani Rustagi[2], Saumya Choudhary[3, 4], Avinash Singh[1*]

[1] Warner College of Dairy Technology , Sam Higginbottom University of Agriculture, Technology & Sciences, Allahabad, India – 211007; [2] Amity Institute of Food Technology, Amity University, Noida, India – 201313; [3] Biomedical Informatics Centre, National Institute of Pathology, ICMR, New Delhi, India – 110029; [4] Dept. of Molecular & Cellular Engineering, Sam Higginbottom University of Agriculture, Technology & Sciences, Allahabad, India 211007
*E-mail: # khan.sheeba778@gmail.com , *avinash.singh@shiats.edu.in*

Peptidomics is an emerging field branching from proteomics that targets endogenously produced protein fragments. The nutritional and functional properties of food proteins have been investigated for many years. The nutritional quality of a protein depends on its amino acid content and on the

physiological utilization of specific amino acids after digestion and absorption. Bioactive peptides are inactive within the sequence of the parent protein and can be released by proteolytic enzymes during gastrointestinal digestion or during food processing, once liberated bioactive peptides can affect numerous physiological functions of the organism. Therapeutic potential for treatment or prevention of disease, bioactive peptides may be used as components in functional foods or nutraceuticals. The main group of the antihypertensive peptides corresponds to the inhibitors of angiotensin converting enzyme (ACE). Synthetic ACE inhibitors like Captopril and lisinopril used to treat hypertension, congestive heart failure, and myocardial infarction but these synthetic ACE inhibitors are known to have strong side effects (Antonios & Macgregor 1995). In contrast, the ACE inhibitory peptides derived from food proteins have not yet shown these side effects. Due to the slow experimental route, bioinformatics was introduced as the solution to quickly mine for bioactive peptides in food. BIOPEP program (Dziuba *et al.,* 2006) has strongly outlined the possible scheme for a peptidomic experiment combining chemometrics and bioinformatics. Chemo-metrical analysis is sufficient to extract chromatographic, electrophoretic or mass spectrometric profile. However, Mass spectrometry is the most suitable method to identify the type of peptide in food hydrolysates. Sequence may be used for database searching to find known activity of the peptide identified. Sequence may also serve as a basis for prediction of the biological or functional properties of peptides using Quantitative Structure-Activity Relationship (QSAR) or another computational approach (Mickiewicz *et al.,* 2008).

References

[1] Minkiewicz P, Dziuba J, Darewicz M, Iwaniak A, Dziuba M, Nałecz D. Food peptidomics. Food technology & biotechnology. 2008; 46(1).

[2] Antonios TF, MacGregor GA. Angiotensin converting enzyme inhibitors in hypertension: potential problems. Journal of Hypertension. 1995; 13:11-16.

[3] Dziuba M, Dziuba J, Minkiewicz P. Design of food protein proteolysis with a view to obtaining bioactive peptides. Polish Journal of Natural Sciences. 2006; 21: 999-1020.

P-77

"CandBase" A Comprehensive Database Covering fifteen classified species of Genus *Candida*

Soumya Sharma[1], Sonali Mishra[2], Krishna Misra[*]

[1]*Indian Institute of Information Technology*

Candida a genus of yeasts and a major human fungal pathogen is known to cause both mucosal and deep tissue infections by colonizing the mucosal surfaces of all humans. More than 200 *Candida* species have been discovered till date in all pathogenic, non-pathogenic and opportunistic states of classification. *Candida albicans* the most popular type of *candida* species, have emerged as opportunistic in nature causing several infections in humans, most common among them being candidiasis. Recently many such *candida* species that were earlier not considered to be pathogenic or opportunistic have emerged so, like *C.glabrata, C.parapsilosis, C.krusei, C. auris* etc. Reports have shown that one of the major causes of *candida* causing pathogenicity and drug resistance is hyphal growth and biofilm formation respectively.

The phylogeny and genomics of these fungal species keeps on modifying, therefore CandBase can be a milestone to incorporate genome and nucleotide information of multiple *Candida* species along with their specific genes responsible for pathogenicity. This can be helpful for the studies related to drug resistance. Analytical tools like BLAST, BCEPRED, ABCPRED and GenSAS have also been incorporated. All 15-selected species, have been categorized and classification of genes has been provided. All sequence files like the nucleotide, gene and genome have been taken from NCBI website and Candida genome database.

In conclusion, recent genomic technological advances have led to a flood of information that pose a crucial challenge of organizing the data that is scattered and ease of availability for researchers. Therefore, a platform where data regarding maximum *Candida* species is available and that can help reduce the time spent in collecting genomic data as well as incorporate and visualize the data with the help of existing web based software's can be of great importance. The classified information

will be helpful for drug resistance studies and antifungal discovery as well.

P-78

Exploring the role of water molecules at ligand binding domain of PDE4B inhibitors: Virtual screening for the development of drugs to treat COPD

Sushil Kumar Kashaw[#*], Priya Singh and Shivangi Agarwal

Department of Pharmaceutical Sciences, Dr. Harisingh Gour University (A Central University), Sagar (MP), India
sushilkashaw@gmail.com[#], priyasingh6734@gmail.com, shivangi.agarwal800@gmail.com*

Chronic obstructive pulmonary disease (COPD) is a respiratory condition associated with emphysema and chronic bronchitis. The potential of PDE4 inhibitors for the treatment of COPD and asthma is well documented[1-2]. Molecular docking is an in-silico method to estimate and study the drug receptor binding pattern. The objective of the study was to explore the binding modes of PDE4B inhibitors with respective receptors and to analyze the role of water molecules present at the active cavity. GOLD and Autodock software were used in the present study. We have performed the docking after deleting all the water molecules from the crystallized structure of PDE4B and it was observed that no hydrogen bonding was formed between the co-crystallized ligand and PDE4B. Similar pattern were observed with all other active ligands (Telafentrine etc). These ligands bind to L393, M-431 and F-446 amino acid residues. However, when the docking was again performed under similar condition with water molecules in the binding site, it was found that the co-crystallize ligand and Telafentrine showed hydrogen bond interactions with Q443 and H234 amino acid residues and with HOH-28 (binding energies of -10.7) and HOH-1008 (binding energies of -10.64). It may be concluded that water molecules play significant role in hydrogen bond interactions and facilitated the interaction of ligand with the receptor.

References

[1] Sundar IK, Nevid MZ, Friedman AE, Rahman I. Cigarette Smoke Induces Distinct Histone Modi fi cations in Lung Cells: Implications for the Pathogenesis of COPD and Lung Cancer. J proteome res 2014; 13: 982-996.

[2] Press NJ, Taylor RJ, Fullerton JD, et al. Discovery and Optimization of 4-(8-(3-Fluorophenyl)-1,7-naphthyridin-6-yl)transcyclohexanecarboxylic Acid, an Improved PDE4 Inhibitor for the Treatment of Chronic Obstructive Pulmonary Disease (COPD). *J Med Chem*. 2015; 58: 6747-6752.

P-79

Comparative bioinformatic analysis of Camel milk Proteins and their Bioactive peptides with potential Anticancer and Antimicrobial activities.

Kumar Udit Saumya[#], Aastha Mittal, Shaswat Singh, Manohar Lal, Neelam Mahala, Atish Paul, Uma S. Dubey[*]

Birla Institute of Technology and Science Pilani, Pilani, Rajasthan-333031, India
E-mail: Kumar Udit Saumya# - uditsaum1992@gmail.com;
Aastha Mittal- aasthamittal14121992@gmail.com;
Shaswat Singh- h2015199@pilani.bits-pilani.ac.in;
Manohar Lal- manohar92826@gmail.com;
Neelam Mahala- p2015419@pilani.bits-pilani.ac.in;
Atish T. Paul- atish.paul@pilani.bits-pilani.ac.in;
Uma S. Dubey - uma@pilani.bits-pilani.ac.in*

Milk proteins are a major source of bioactive peptides which stand as potential biomolecules in therapeutics against several diseases(1) such as Cancer, Neurological, Autoimmune, and now as bactericidal agents too. In the present study, camel milk's Alpha-lactalbumin and six other major proteins have been studied, to determine the formers ability of forming a BAMLET or HAMLET like tumoricidal complex with oleic acid (2,3), and the latter to possess any anticancer or antimicrobial peptides. Insilico protein hydrolysis with 28 different enzymes was carried out with BIOPEP's Enzyme Action tool, while AntiCp and CAMP-R3 was used to screen the generated peptides for their anticancer and antimicrobial properties. Both AntiCP and CAMP-R3 use multiple algorithms for higher accuracy in prediction such as Support Vector Machine (SVM), Random Forest (RF), Artificial Neural Networks (ANN), Discriminant

Analysis (DA) and Binary profiles based models. Furthermore the characterization and comparison with already experimentally validated peptides using parameters such as Amino acid composition, Molecular weight, pI, GRAVY index, Boman index, Aliphatic index and Instability index was done. Also, comparison has been made between the primary, secondary and tertiary structure followed by binding efficiency of alphalactalbumin with oleic acid using AutoDock tools. This study highlights a very strong sequential, structural homology based on which we propose the presence of an anticancer and antimicrobial molecule in camel milk.

References

[1] N. Škalko-Basnet, Biologics: Targets & Therapy, 8:107-114 (2014).

[2] Catharina Svanborg et.al, HAMLET kills tumor cells by an apoptosis like mechanism- cellular, molecular and therapeutic aspects: Advances in cancer research, 88: 1-29 (2005).

[3] Paul Rammer *et.al,* BAMLET activates lysosomal cell death program in cancer cells: Molecular Cancer Therapeutics, 9: (2010).

P-80

Attraction between like charged proteins: A theoretical study of bovine β-lactoglobulin dimer

Rakesh Srivastava[#], Pradipta Bandyopadhyay*

[#,] Jawaharlal Nehru University, New Delhi, India*
*E-mail: # allahabad.21@gmail.com * praban07@gmail.com*

Like charge attraction under the effect of counterions is known to be important in biological systems *e.g*. like charged protein subunits binding, DNA condensation *etc* [1]. The native bovine β-lactoglobulin (β-LG) monomer, which has +13 units of positive charge at pH 3.0, exists as major species under low salt concentrations. It has been proved that an increase in the salt concentration stabilizes the dimer. Experimentally it has been shown that at pH 3.0, temperature 20° C and 1.0 mg mL^{-1} concentration of β-LG, more than 80 % (w/w) of β-LG forms dimer in the Presence of 1 M NaCl and dimer formation is 50 % at 0.1 M NaCl [2]. 3D-RISM (three dimensional reference interaction site model), which is a statistical mechanics based theory, has been shown to produce solvation thermodynamic properties in good agreement with experiments [3]. In this work, with the help of 3D-RISM theory, we have studied the binding free energy variation of dimer formation of β-LG with varying salt (NaCl) concentration. Our results also show agreement to experimental results.

References

[1] Bloomfield, Victor A. "DNA condensation by multivalent cations." Biopolymers 44.3 (1997): 269-282.

[2] Sakurai, Kazumasa, Motohisa Oobatake, and Yuji Goto. "Salt dependent monomer–dimer equilibrium of bovine â lactoglobulin at pH 3." Protein Science 10.11 (2001): 2325-2335.

[3] Imai, Takashi, Andriy Kovalenko, and Fumio Hirata. "Solvation thermodynamics of protein studied by the 3D-RISM theory." Chemical Physics Letters 395.1 (2004): 1-6.

P- 81

A computational approach towards targeting protein-protein interaction by alkaloids: A futuristic therapeutic intervention strategy for breast cancer impediment

Sameeksha Tiwari[#], Manika Awasthi, Swati Singh, Veda P. Pandey, Upendra N. Dwivedi*

Department of Biochemistry, Centre of Excellence in Bioinformatics, Bioinformatics Infrastructure Facility, University of Lucknow, Lucknow-226007, U.P., India.
[#]presenting author
**Corresponding author: E-mail: upendradwivedi@hotmail.com*

Protein-protein interactions (PPI) exist as potential intervention points for drug against several diseases and can be a new emerging class of novel therapeutic targets. With the availability of a number of databases, focusing on PPI complexes and their interaction network has provided a huge resource for such kind of analyses. Keeping this in view the present study was initiated to analyze interaction of tumour suppressor protein p53 (TP53) and breast cancer associated protein (BRCA1) as novel and promising target against breast cancer. Using computational approaches such as protein-protein docking, hot spot analyses, molecular docking and molecular dynamics simulation (MDS), a methodology was set for stepwise analyses

of the interaction of the wild type and mutant TP53 with that of wild type BRCA1 and their modulation by alkaloids. Protein-protein docking method was used to generate both wild type and mutant complexes of TP53-BRCA1. Subsequently, the complexes were docked using sixteen different alkaloids, fulfilling ADMET and Lipinski's rule of five criteria, and were compared with that of a well-known inhibitor of PPI, namely nutlin. The alkaloid dicentrine was found to be the best docked alkaloid among all the docked alklaloids as well as that of nutlin. Furthermore, MDS analyses of both wild type and mutant complexes with the best docked alkaloid i.e. dicentrine, revealed higher stability of mutant complex than that of the wild one, in terms of average RMSD, RMSF and binding free energy, corroborating the results of docking. Results suggested more pronounced interaction of BRCA1 with mutant TP53 leading to increased expression of mutated TP53 thus showing a dominant negative gain of function and hampering wild type TP53 function leading to tumour progression.

References

[1] Modell A.E., Blossr S.L., Arora P.S. (2016). Systematic targeting of protein-protein interactions. Trends Pharmacol Sci, 20.

[2] Singh S., Das T., Awasthi M., Pandey V. P., Pandey B., Dwivedi U.N. (2015). DNA topoisomerase-directed anticancerous alkaloids: ADMET-based screening, molecular docking, and dynamics simulation. Biotechnol Appl Biochem, DOI: 10.1002/bab.1346.

[3] Rasti M. and Azimi T. (2015). Tp53 binding to BRCA1 and RAD51 in MCF7 and MDA-MB-468 Breast Cancer Cell lines In vivo and In Vitro. Avicenna J Med Biotech, 7, 76-79.

P-82

Bioinformatics intervention in cotton (*Gossypium hirsutum* L.) transcriptome analysis

Praveen Prajapat, Diwakar Singh[#*] and K. P. Suthar

Department of Plant Molecular Biology and Biotechnology, ASPEE College of Horticulture and Forestry, Navsari Agricultural University, Navsari, Gujarat – 396 450
**drdiwakarbiochem@gmail.com*

Cotton is one of the most important economic crops in the world due to its natural textile fiber and drought stress leads to penalty for its economic yield. Drought stress leads to several changes at transcriptomic, proteomic and metabolomics level. Identification of differentially regulated transcripts may discover novel transcription factors, genes and biochemical pathways that impart tolerance against drought stress and bioinformatics tools may gear up this identification process. Transcriptome study was carried out in cotton under drought stress condition using illumina Hiseq 2500 sequencer to identify differentially regulated transcripts. RNA samples were extracted from leaves of two cotton genotypes namely Gcot-16 (drought tolerant) and GBHV-177 (drought susceptible) under drought stress condition. The raw sequences were trimmed using pearl script program and more number of bases was found in drought susceptible genotype than tolerant. Pre-processed reads were aligned to the available *Gossypium hirsutum* L. (AD_1) genome and gene model were downloaded from Cotton Genome Project. The alignment was performed using Tophat program (version 2.1.1). A total of 48,57,15,286 pooled high quality reads of all the cotton leaf sample were mapped on the reference genome *Gossypium hirsutum* L., having 24,285.77 MB bases. Out of the total reads, 42,65,93,714 reads (91.97%) were mapped to reference genome of *Gossypium hirsutum* L. and 3,90,61,718 (8.03%) reads remained unmapped. Gene expression estimation was performed using Cufflinks program and differential gene expression analysis was done with Cuffdiff program. A complete linkage hierarchical analysis was done on top differentially expressed genes obtained from NOISeq using multiple experiment viewer (MEV v4.8.1). Ven diagram and volcano plots showed that more number of genes was up regulated in drought tolerant Gcot-16 genotype than drought susceptible GBHV-177 genotype. Heat map was also constructed using the log transformed and normalized value of genes based on Pearson's un-centered correlation distance as well as based on complete linkage method. GO assignments clarified the functions of the up regulated transcripts in tolerant and susceptible genotypes and KEGG pathway analysis showed that genes associated with starch & sucrose metabolism, phenylpropanoid biosynthesis, amino sugar and nucleotide metabolism, glycerophospholipid metabolism were highly enriched in tolerant genotype than susceptible. Whereas, genes related to fatty acid biosynthesis was highly enriched in susceptible genotype. By

using different bioinformatics tools several genes, transcription factors and biochemical pathways could be identified that may be involved in drought tolerant mechanism in cotton.

P-83

Harnessing Big Data for Chemical Toxicity Profiles

Rohit Bhatia*

People for Ethical Treatment of Animal, PO Box 28260
Juhu, Mumbai, India
E-mail: RohitB@petaindia.org

Historically, tests on animals have been used in an attempt to assess the toxicity of chemicals. However, these tests are expensive and time-consuming and are associated with significant animal-welfare and other ethical concerns because of the pain and suffering experienced by the animals. Furthermore, data generated using experiments on animals can rarely be extrapolated to humans, as the differences in anatomy, biochemistry, and physiology between animal species are vast.

To overcome the drawbacks associated with using animals to predict the toxic effects of chemicals, forward-thinking scientists are developing and validating animal-free methods that use human cells and tissues, high-throughput experiments, and *in silico* models. The large quantities of data generated by these techniques provide unprecedented opportunities for predicting chemical toxicity. Although many conventional database management tools and data processing applications are not sophisticated and powerful enough to handle these large amounts of complex data, recent advances in data science and big data analytics offer new methods for data-driven predictions of chemical toxicity profiles. In this presentation, an overview of the big data modelling and mining tools that can be used to predict toxicity profiles will be given.

P-84

Qsar Ofpde4 Inhibitors for Potential Therapy Against Chronic Obstructive Pulmonary Disease (COPD)

Dr Ratnesh Das

Department of Chemistry
Dr Hari singh Gour Central University Sagar

Chronic obstructive pulmonary disease is the long term gulp of polluted air, smoking of cigarette, chemical and environmental dusts. They are also causes the inflammation in lungs. COPD is collection of lung diseases including chronic bronchitis, emphysema and chronic obstructive airways disease.PDE4 inhibitor prevents the hydrolysis of cyclic adenosine monophosphate. PDE4 enzyme is present in inflammatory cells.Selective PDE4 inhibitor has antiinflammatory effect such as inhibition cytokinine and inflammatory cells such as neutrophils esinophil or microphages and T cells. The standard drugs of PDE4 inhibitor Roflumilast, and cilomilast are in clinical trial. But the major side effects of these drugs are severe nausea and vomiting. The developmentof a new drug by the help Quantitative Structure ActivityRelationshipforselective PDE4 inhibitor . A literature survey of reported Triazines series used for drug design. The development of 2D and 3D QSAR model using different methods like(PLS)Partial Least Square and (kNN) k Nearest Neighbour were performed. The value of r^2 (0.7871) q^2 (0.7146) and pred.R^2 (0.6225) that showed predicted activity of 78% and 62% of tanning set and test set. Selected descriptors for the PLS model were various physicochemical descriptors while, hydrophobic, steric and electrostatic descriptors were considered for kNN model. In COPD the QSAR study showed that a new series of the Triazines derivatives compounds having best PDE4 inhibitor properties.

Keywords: COPD, Triazines, QSAR studies. PLS, kNN, PDE 4 Inhibitor.

P- 85

Molecular level insights into the effect of D59P mutation on the aggregation propensity of β_2-microglobulin

Simranjeet Singh Narang, Suniba Shuaib, Deepti Goyal and Bhupesh Goyal#*

Department of Chemistry, School of Basic and Applied Sciences, Sri Guru Granth Sahib World University, Fatehgarh Sahib–140406, Punjab, India
E-mail: bhupesh@iitbombay.org

The accumulation of amyloidogenic β_2–microglobulin (â$_2$m) protein around skeletal joints and bones lead to a severe condition known as Dialysis-related amyloidosis (DRA). A critical role of the DE loop region for β_2m stability and amyloid aggregation propensity has been highlighted in the recent studies.[1,2] However, the molecular mechanism of enhanced β_2m aggregation due to D59P mutation in the DE loop region remains elusive. In this regard, explicit-solvent molecular dynamics (MD) simulations were performed in the present study to elucidate the key structural and dynamic changes in the wild type (wt) β_2m upon D59P mutation.[3] MD simulations reveal a decrease in the average number of hydrogen bonds in the loop regions of â$_2$m on D59P mutation that enhances conformational flexibility, which in turn leads to higher aggregation propensity of D59P as compare to wt β_2m. D59P cover a larger region of phase space and display a higher trace value than wt β_2m, which suggest an overall enhancement in the conformational flexibility. D59P display two minimum energy basins in the free energy landscape that are associated with thermodynamically less stable conformational states as compare to single minimum energy basin in wt β_2m. The present study provides atomic level details into the molecular mechanism behind the higher aggregation propensity of D59P as compare to wt β_2m.

Keywords: β_2–microglobulin (β_2m); D59P mutation; DE loop; molecular dynamics; amyloid aggregation

References

[1] Camilloni, C.; Sala, B. M.; Sormanni, P.; Porcari, R.; Corazza, A.; De Rosa, M.; Zanini, S.; Barbiroli, A.; Esposito, G.; Bolognesi, M.; Bellotti, V.; Vendruscolo, M.; Ricagno, S. *Sci. Rep.* 2016, *6*, 25559.

[2] Natalello, A.; Relini, A.; Penco, A.; Halabelian, L.; Bolognesi, M.; Doglia, S. M.; Ricagno, S. *PLoS One* 2015, *10*, e0122449.

[3] Narang, S. S.; Shuaib, S.; Goyal, D.; Goyal, B. *J. Cell. Biochem.* 2017, DOI: 10.1002/jcb.26241.

P-86

Role of Local and Nonlocal Interactions and Site-directed Point Mutations in Folding and Misfolding of Globular Proteins

Adesh Kumar#, Anupaul Baruah and Parbati Biswas*

Department of Chemistry, University of Delhi Delhi-110007
**E-mail: pbiswas@chemistry.du.ac.in*

A Monte Carlo simulation based sequence design method[1] is proposed to study the role of the local and the nonlocal interactions[2] and site-directed point mutations[3] in protein folding, misfolding and unfolding. A statistical potential is developed from the compilation of a data set of proteins, which accounts for the respective contribution of the local and the nonlocal interactions. Sequences are designed through a combination of positive and negative design by a Monte Carlo simulation in the sequence space. The weights of the local and the nonlocal interactions are tuned appropriately to study the role of the local and the nonlocal interactions in the folding, misfolding and unfolding of the designed sequences. The site-directed point mutation procedure is developed and applied on the designed sequences of real proteins to generate a diverse set of mutated sequences. A clash and match procedure is proposed, which may be used to predict the number of residue pairs in a sequence with unfavourable and favourable interactions, respectively. Results suggest that the nonlocal interactions are the primary determinant of protein folding while the local interactions may be required but not always necessary. The nonlocal interactions mainly guide the polypeptide chain to form compact structures but do not differentiate between the native-like conformations, while the local interactions stabilize the target conformation against the native-like

competing conformations. The number of clashing and matching residue pairs may indicate whether the mutated sequence would be folded or misfolded. The study concludes that the local interactions govern the fold-misfold transition, while the nonlocal interactions regulate the fold-unfold transition of proteins. However, for proteins with predominantly β-sheet content the nonlocal interactions control both fold-misfold and fold-unfold transitions.

References

[1] A. Kumar, A. Baruah and P. Biswas, *J. Chem. Phys.*, 146, 065102 (2017).

[2] K. A. Dill, *Biochemistry*, 29, 7133 (1990).

[3] A. Baruah and P. Biswas, *Phys. Chem. Chem. Phys.*, 16, 13964 (2014).