

IWBBIO 2018

INTERNATIONAL WORK-CONFERENCE ON
BIOINFORMATICS AND
BIOMEDICAL ENGINEERING

PROCEEDINGS EXTENDED ABSTRACTS

25-27 April, 2018

IWBBIO 2018
International work-conference on
Bioinformatics and biomedical engineering

Proceedings Extended abstracts
25-27 April,2018
Granada (SPAIN)

Authors .

Daniel Castillo

Juan Manuel Gálvez

Maria José Sáez

Fernando Rojas

Luis Javier Herrera

Ignacio Rojas

I.S.B.N: 978-84-17293-36-9

Legal Deposit: Gr- 605-2018

Edit and print : Editorial Godel S.L.

All rights reserved to the authors . Is strictly prohibited
without written permission of the copyright holders under sanctions
established by law , the total or partial reproduction of this work.

We are proud to present the set of abstract for the fourth edition of the IWBBIO conference "International Work-Conference on Bioinformatics and Biomedical Engineering" held in Granada (Spain) during April 25-27, 2018.

The IWBBIO 2018 (International Work-Conference on Bioinformatics and Biomedical Engineering) seeks to provide a discussion forum for scientists, engineers, educators and students about the latest ideas and realizations in the foundations, theory, models and applications for interdisciplinary and multidisciplinary research encompassing disciplines of computer science, mathematics, statistics, biology, bioinformatics, and biomedicine.

The aims of IWBBIO 2018 is to create a friendly environment that could lead to the establishment or strengthening of scientific collaborations and exchanges among attendees, and therefore, IWBBIO 2018 solicited high-quality original research papers (including significant work-in-progress) on any aspect of Bioinformatics, Biomedicine and Biomedical Engineering.

New computational techniques and methods in machine learning; data mining; text analysis; pattern recognition; data integration; genomics and evolution; next generation sequencing data; protein and RNA structure; protein function and proteomics; medical informatics and translational bioinformatics; computational systems biology; modelling and simulation and their application in life science domain, biomedicine and biomedical engineering were especially encouraged. The list of topics in the successive Call for Papers has also evolved, resulting in the following list for the present edition:

- Computational proteomics: Analysis of protein-protein interactions. Protein structure modelling. Analysis of protein functionality. Quantitative proteomics and PTMs. Clinical proteomics. Protein annotation. Data mining in proteomics.
- Next generation sequencing and sequence analysis: De novo sequencing, re-sequencing and assembly. Expression estimation. Alternative splicing discovery. Pathway Analysis. Chip-seq and RNA-Seq analysis. Metagenomics. SNPs prediction.
- High performance in Bioinformatics. Parallelization for biomedical analysis. Biomedical and biological databases. Data mining and biological text processing. Large scale biomedical data integration. Biological and medical ontologies. Novel architecture and technologies (GPU, P2P, Grid,...) for Bioinformatics.
- Biomedicine. Biomedical Computing. Personalized medicine. Nanomedicine. Medical education. Collaborative medicine. Biomedical signal analysis. Biomedicine in industry and society. Electrotherapy and radiotherapy.
- Biomedical Engineering. EComputer-assisted surgery. Therapeutic engineering. Interactive 3D modelling. Clinical engineering. Telemedicine. Biosensors and data acquisition. Intelligent instrumentation. Patient Monitoring. Biomedical robotics. Bio-nanotechnology. Genetic engineering.
- Computational systems for modelling biological processes. Inference of biological networks. Machine learning in Bioinformatics. Classification for biomedical data. Microarray Data Analysis. Simulation and visualization of biological systems. Molecular evolution and phylogenetic modelling.

- Healthcare and diseases. Computational support for clinical decisions. Image visualization and signal analysis. Disease control and diagnosis. Genome-phenome analysis. Biomarker identification. Drug design. Computational immunology.
- E-Health. E-Health technology and devices. E-Health information processing. Telemedicine/E-Health application and services. Medical Image Processing. Video techniques for medical images. Integration of classical medicine and E-Health.

During IWBBIO 2018 several Special Sessions will be carried out. Special Sessions will be a very useful tool in order to complement the regular program with new and emerging topics of particular interest for the participating community. Special Sessions that emphasize on multi-disciplinary and transversal aspects, as well as cutting-edge topics are especially encouraged and welcome, and in this edition of IWBBIO 2018 are the following:

Special Session IWBBIO 2018

SS1. Generation, Management and Biological Insights from Big Data.

Organizer: Dr. Anagha Joshi, Group leader in the Division of Developmental Biology at the Roslin Institute, University of Edinburgh.

Website: <https://www.ed.ac.uk/roslin/about/contact-us/staff/anagha-joshi>

SS2. Challenges in smart and wearable sensor design for mobile health.

Organizers: Prof. Dr. Natividad Martínez Madrid. Head of the Internet of Things laboratory and Director of the AAL-Living Lab at Reutlingen University. Reutlingen University. Department of Computer Science, Alteburgstr. 150, D-72762 Reutlingen (Germany).

Prof. Dr. Juan Antonio Ortega. Director of the Centre of Computer Scientific in Andalusia (Spain) www.cica.es and the head of the research group IDINFOR (TIC223). University of Sevilla, ETS Ingeniería Informática. Spain.

Prof. Dr. Ralf Seepold. Head of the Ubiquitous Computing Lab at HTWG Konstanz. Department of Computer Science, Brauneeggerstr. 55, D-78462 Konstanz (Germany)

Websites: <http://iotlab.reutlingen-university.de>; <http://madeirasic.us.es/idinfor/>; <http://uc-lab.in.htwg-konstanz.de>

SS3. Challenges and advances in measurement and self-parametrization of complex biological systems

Organizer: Dipl.-Ing. Jan Urban, Ph.D.. Head of laboratory of signal and image processing. University of South Bohemia in České Budějovice. Faculty of Fisheries and Protection of Waters. South Bohemian Research Center of Aquaculture and Biodiversity of Hydrocenoses. Institute of Complex Systems. Czech Republic.

Websites: www.frov.jcu.cz/en/institute-complex-systems/lab-signal-image-processing

SS4. High-throughput bioinformatic tools for medical genomics.

Organizers: Prof. M. Gonzalo Claros, Department of Molecular Biology and Biochemistry, University of Málaga, Spain.

Dr. Javier Pérez Florido, Bioinformatics Research Area, Fundación Progreso y Salud, Seville, Spain.

SS5. Drug Delivery System Design Aided by Mathematical Modelling and Experiments

Organizers: Ph.D cand Kristinn Gudnason. Faculty of Industrial Engineering, Mechanical Engineering and Computer Science, University of Iceland, Iceland.

Prof. Fjola Jonsdottir. Faculty of Industrial Engineering, Mechanical Engineering and Computer Science, University of Iceland, Iceland.

<p>Prof. Emeritus Sven Sigurdsson. Faculty of Industrial Engineering, Mechanical Engineering and Computer Science, University of Iceland, Iceland.</p> <p>Prof. Mar Masson. Faculty of Pharmaceutical Science, University of Iceland, Iceland.</p>
<p>SS6. Molecular studies on inorganic nanomaterials for therapeutical and diagnosis applications</p> <p>Organizers: Dr. C. Ignacio Sainz-Díaz. Instituto Andaluz de Ciencias de la Tierra, CSIC/UGR, Granada, Spain. Expert in organic chemistry and interactions of bioactive compounds on mineral surfaces by computational chemistry with experience in I+D of pharmaceutical industry.</p> <p>Dr. Carola Aguzzi. Dpto. Tecnología Farmacéutica. Universidad de Granada, Granada, Spain. Expert in pharmaceutical technology and applications of natural inorganic excipients.</p>
<p>SS7. Little-big data. Reducing the complexity and facing uncertainty of highly under-determined phenotype prediction problems.</p> <p>Organizer: Prof. Juan Luis Fernández-Martínez. Mathematics Department. Applied Mathematics Section. Director of the Group of Inverse Problems, Optimization and Machine Learning. University of Oviedo. Spain.</p> <p>Website: https://www.researchgate.net/profile/Juan_Luis_Fernandez-Martinez</p>
<p>SS8. Interpretable models in biomedicine and bioinformatics</p> <p>Organizers: Prof. Alfredo Vellido Intelligent Data Science and Artificial Intelligence (IDEAI) Research Center. Universitat Politècnica de Catalunya, Barcelona, Spain.</p> <p>Prof. Sandra Ortega-Martorell Department of Applied Mathematics, Liverpool John Moores University, Liverpool, UK.</p> <p>Prof. Alessandra Tosi Mind Foundry Ltd., Oxford, UK</p> <p>Prof. Iván Olier Caparoso MMU Machine Learning Research Lab, Manchester Metropolitan University, Manchester, UK</p>
<p>SS9. Medical Planning: Management System for Liquid Radioactive Waste in Hospital Design.</p> <p>Organizer: Dr. Khaled El-Sayed is presently an assistant professor of Biomedical Engineering with the department of Electrical and Medical Engineering at Benha University, Egypt.</p>
<p>SS10. Bioinformatics tools to integrate omics dataset and address biological questions</p> <p>Organizer: Dr. Domenica Scumaci, PhD Laboratory of Proteomics, Dpt. of Experimental and Clinical Medicine; Magna Græcia University of Catanzaro, "S.Venuta" University Campus, viale S.Venuta, 88100 Catanzaro</p>
<p>SS11. Understanding the Mechanisms of Variant Effects on Human Disease Phenotype.</p> <p>Organizer: Anna Panchenko, PhD. Head, Computational Biophysics Group Computational Biology Branch, National Center for Biotechnology Information, National Institutes of Health, Bethesda, USA</p> <p>Website: https://www.ncbi.nlm.nih.gov/CBBResearch/Panchenko/</p>

In this edition of IWBBIO, we are honored to have the following invited speakers:

- **Prof. Joaquín Dopazo**, Fundación Progreso y Salud, Clinical Bioinformatics Research Area, Sevilla, Spain.
- **Prof. Luis Rueda**, Professor, School of Computer Science, Pattern Recognition and Bioinformatics Lab, Windsor Cancer Research Group, University of Windsor.
- **Dr. Anagha Joshi**, Bioinformatics Group Leader, Developmental Biology Division, The Roslin Institute, University of Edinburgh, UK.
- **Prof. FangXiang Wu**, Prof. FangXiang Wu Ph.D., P.Eng, SMIEEE Professor, Division of Biomedical Engineering, Professor, Department of Mechanical Engineering, College

of Engineering, University of Saskatchewan, 57 Campus Dr., Saskatoon, SK Canada, S7N5A9.

- **Prof. Jiayin Wang**, Professor, Department of Computer Science and Technology, Xi'an Jiaotong University (China).

We wish to thank to our main organizer/institutions, Dept. Computer Architecture & Computer Technology, Faculty of Science and CITIC-UGR from the University of Granada and BioMed Central for their support and grants.

We wish also to thank to the Editor-in-Chief of different international journals for their interest in editing special issues from the best papers of IWBBIO 2018. We would also like to express our gratitude to the members of the different committees for their support, collaboration and good work. Especially to the organizers of the Special Sessions, thank a lot for the excellent ideas/topics/subject for the special sessions and effort carried out !! Thanks to the invited speakers for their effort and interest in the realization of these excellent presentations. Thanks to all reviewers for selfless effort reviewing contributions, which positively influences the quality of the final papers presented at the conference.

Last, but not least, thanks to all the participant of IWBBIO.

Granada, April 2018,

Daniel Castillo
Juan Manuel Gálvez
María José Sáez
Fernando Rojas
Luis Javier Herrera
Ignacio Rojas

IWBBIO 2018 EXTENDED ABSTRACTS

Bioinformatics for healthcare and diseases

PlasmidID: a mapping based tool for plasmid identification, annotation, classification and representation. Application as high-risk strains plasmid detection.	1
<i>Pedro J. Sola-Campoy, Sara Monzn, Mara Prez-Vzquez, Beln Aracil, Jos Campos, Jess Oteo and Isabel Cuesta</i>	
DB2DB: An enrichment databases approach for medical domain	6
<i>Zina Nakhla, Kaouther Nouira and Ahmed Ferchichi</i>	
The importance of using quality control samples to apply data processing methods in large fingerprinting metabolomic studies. A practical case with 288 samples of systemic autoimmune diseases.	17
<i>lvoro Fernndez Ochoa, Isabel Borrs Linares, Rosa Quirants Pine, Marta E Alarcn Riquelme and Antonio Segura Carretero</i>	
Detecting Cancer-Associated Epistatic Gene Variants in Lung Adenocarcinoma	21
<i>Jaume Sastre Tomas, Jairo Rocha, Alexander Damia Heine Suer and Emidio Capriotti</i>	
GPU-Low-Energy Tracking of the Left Ventricle in the Cloud	25
<i>Sidi Ahmed Mahmoudi, Mohammed Ammar, Mohammed Amin Belarbi and Amine Abbou</i>	
Web based application for accurately classifying cancer type from microarray gene expression data using a support vector machine learning algorithm	33
<i>Shrikant Pawar</i>	
Interaction between <i>Aedes aegypti</i> CPB1 and viral proteins	44
<i>Edgar Lpez Lpez and Carolina Barrientos Salcedo</i>	

Biomedical Engineering

Aging Effect on the Pelvis to Head Attenuation of Upper Body Rotation during Walking .	49
<i>Hyeongmin Jeon, Jaehoon Heo, Euibum Choi and Gwang Moon Eom</i>	
Potentiometric Screen-Printed Sensors for Wireless Monitoring of Wound Healing	52
<i>Andrzej Pepowski, Daniel Janczak, ukasz Grski, Micha Zbie, Dariusz Obrbski and Magorzata Jakubowska</i>	
Graphene layers and coatings in biomedical applications	56
<i>Lucja Dybowska-Sarapuk, Andrzej Kotela and Malgorzata Jakubowska</i>	
Computational Modelling of the Possible Cancer Synergic Treatment Combining Oncolytic Treatment and Induced Immune Cytotoxicity.	61
<i>Jos-Antonio Lpez-Valverde</i>	
Adapting and step by step refinement of the Vegetative Vigour Terrestrial Plant Test for assessing ecotoxicity of aerosol samples	73
<i>Katalin Hubai, Eszter Horvth, Bettina Eck-Varanka, Gbor Teke, dm Tth and Nra Kovts</i>	

Biomedical image analysis

Classification in fMRI studies: in search of brain informative regions.	75
<i>Juan E Arco, Paloma Daz-Gutierrez, Javier Ramrez and Mara Ruz</i>	

Molecular Modeling of The Adsorption Of 5-Aminosalicylic Acid on the Kaolinite and Halloysite Nanotube Surfaces	80
<i>Ana Borrego-Sanchez, Mahmoud E. Awad, Elisabeth Escamilla-Roa, Esperanza Carazo and C. Ignacio Sainz-Daz</i>	
Contrast enhancement methods for images from the light microscope	85
<i>Frantiek Jablonk, Libor Harga, Duan Koniar and Jozef Volk</i>	

Biomedicine

A Novel Demodulation Technique to Estimate Continuous Blood Pressure from Photoplethysmographic Signals	89
<i>Fabian Kern and Stefan Bernhard</i>	
Bioinformatic analysis of selected aptamer sequences allows the identification of RNA tools for the functional analysis of West Nile virus genomic RNA elements	91
<i>Cristina Romero-Lpez, Beatriz Berzal-Herranz and Alfredo Berzal-Herranz</i>	
The effect of c-myc monoclonal antibody for gastric cancer cells in vivo and in vitro	94
<i>Hu Haixia, Su Yanzhuo, Song Bin and Du Juan</i>	

Challenges and advances in measurement and self-parametrization of complex biological systems

CasANN: Targeted Identification of CRISPR Associated Gene Operons from Next Generation Sequences.	114
<i>Corey Hudson, Jerilyn Timlin and Kelly Williams</i>	
Characterization of the telomeric transcriptome from Chironomus sp. Sequence analyses and structural predictions	116
<i>Mercedes Iriarte Cela, Jos Luis Martnez-Guitarte and Mercedes de La Fuente</i>	

Computational genomics and proteomics

Improving the performance of pathway extraction methods by infeasibilities removal	121
<i>Jose F Hidalgo, Francisco Guil and Jos M Garca</i>	
SalivaPRINT Toolkit: Development, Challenges and Applications	137
<i>Igor Cruz, Eduardo Esteves, Mnica Fernandes, Nuno Rosa, Maria Jos Correia and Marlene Barros</i>	
When intrinsically disordered proteins are ordered	139
<i>Erzsbet Fich, Istvn Simon and Blint Mszros</i>	

Computational systems for modelling biological processes

The binding of agonists to the NOP receptor revealed in silico by Multi-flexible docking and QM/MM.	141
<i>Stefano Della Longa and Alessandro Arcovito</i>	
Classification of cardiovascular pathologies in artificial signals of a lumped parameter model using a naive Bayes algorithm	143
<i>Stefan Krickl, Manuel Debic and Stefan Bernhard</i>	

Putative oxidative folding pathways of the disulfide rich inhibitory peptide Tridegin	145
<i>Ajay Abisheck Paul George, Arijit Biswas, Monica Sudarsanam and Diana Imhof</i>	
Red blood cell model with different implementation of viscoelastic parameter	147
<i>Mariana Ondruov</i>	
Development of multi-agent technology for prediction of the "structure-property" dependence of drugs on the basis of modified algorithms of artificial immune systems	152
<i>Samigulina Galina and Samigulina Zarina</i>	

Drug Delivery System Design Aided by Mathematical Modelling and Experiments

Impedance Analysis of Different Shapes of the Normal and Malignant White Blood Cells .	154
<i>Sameh Sherif</i>	
Controlling drug delivery from multi-layer polymeric coated capsules	158
<i>Giuseppe Pontrelli, Elliot Carr, Badr Kaoui, Marco Lauricella and Sauro Succi</i>	
Drug diffusion properties of hydrogels: numerical and experimental studies	161
<i>Svetlana Solodova, Andreia Pimenta, Kristinn Gudnason, Ana Paula Serro, Fjola Jonsdottir, Sven Sigurdsson and Mar Masson</i>	
Numerical evaluation of skin sub layer properties	163
<i>Bergthra Snorraddtir, Kristinn Gunason, Fjla Jnsdttir, Sven Sigurdsson and Mr Msson</i>	
Modelling the drug release from intraocular lens material with variable loading times	165
<i>Ana Topete, Kristinn Gudnason, Benilde Saramago and Ana Paula Serro</i>	

Healthcare and diseases

The prognostic value of clinical variables for the treatment response in rheumatoid arthritis	167
<i>Alexander Platzter, Daniela Sieghart, Farideh Alasti, Guenter Steiner and Josef Smolen</i>	
The Study of Complications of Intestinal Obstruction Catheter	170
<i>Song Bin, Hu Haixia, Du Juan and Su Yanzhuo</i>	
Anatomic Variation of Celiac Axis Influencing The D2 Radical Resection For Distal Gastric Cancer	175
<i>Song Bin, Hu Haixia, Du Juan and Su Yanzhuo</i>	
Clinical study on the treatment of metastatic malignant bowel obstruction with transgastric Intestinal Obstruction Catheter arrangement small intestinal enterostomy	179
<i>Song Bin, Hu Haixia, Du Juan and Su Yanzhuo</i>	
Equilibrium and Thermodynamic Studies of Isoniazid Adsorption by Montmorillonite	184
<i>Esperanza Carazo, Ana Borrego-Snchez, Ftima Garca-Villn, Rita Snchez-Espejo, Pilar Cerezo, Carola Aguzzi and Csar Viseras</i>	
The PERSON project: A brain computer interface serious game for neurorehabilitation . . .	189
<i>Alfonso Monaco, Nicola Amoroso, Roberto Bellotti, Paolo Da Pelo, Domenico Diacono, Gianluca Sforza and Sabina Tangaro</i>	
Cytotoxicity of Secnidazole cocristals on cells infected with Trypanosoma cruzi	192
<i>Carolina Barrientos-Salcedo, Micheel Merari Vichi Ramirez, Edgar Lpez Lpez, Jorge Guillermo Dominguez Chavez, Karina Mondragon Vzquez, Misaela Francisco Mrquez, Ignacio Sainz-Daz, Catalina Soriano-Correa, Bertha Josefina Espinoza Gutierrez and Ignacio Martinez Martinez</i>	

Experimental Design for Optimizing the Properties An-timicrobial Chitosan Derivatives and Investigation of the Structure Activity Relationship	194
<i>P. Sahariah, B.S. Snorraddtir, Martha . Hjlmarsdttir, lafur E. Sigurjnsson and M. Msson</i>	

High-throughput bioinformatic tools for medical genomics

Design of genus-specific probes and primers for detection and identification of viral DNA in environmental samples using next-generation sequencing method	196
<i>Andrey Ayginin, Anna Speranskaya, Alina Matsvai, Vladimir Dedkov, Marina Safonova, Ekaterina Pimkina, Dmitry Shagin, Mikhail Markelov, German Shipulin and Kamil Khafizov</i>	
PIKAVIRUS: A Metagenomics Tool for Viral Community Analysis	207
<i>Andrea Rubio, Sara Monzon, Pedro J. Sola-Campoy and Isabel Cuesta</i>	
Gene Expression-based Cancer Classification with Handling the Class Imbalance Problem ..	213
<i>Emad Ramadan and Sadam Al-Azani</i>	

Next generation sequencing and sequence analysis

Evaluation of differential gene expression workflow variability through RNA-Seq experiments.	228
<i>Ivaro Prez Sala and Ignacio M. Larryoz Roldn</i>	
New features and recipes for simulating structural mutations with BAMSurgeon	231
<i>Adam Ewing</i>	
The roles and mechanisms of celecoxib in gastric cancer NCI-N87 cells	233
<i>Song Bin, Du Juan and Su Yanzhuo</i>	

PlasmidID: a mapping based tool for plasmid identification, annotation, classification and representation. Application as high-risk strains plasmid detection.

Pedro J. Sola-Campoy^[1,2], Sara Monzón^[2], María Pérez-Vázquez^[1], Belén Aracil^[1], José Campos^[1], Jesús Oteo^[1], Isabel Cuesta^[2]

¹ Reference and Research Laboratory for Antibiotic Resistance, CNM, ISCIII, Majadahonda, Madrid.

² Bioinformatic Unit, ISCIII, Majadahonda, Madrid.

Abstract. Plasmids are responsible for dissemination of antibiotic resistance genes (ARG), virulence factors and other accessory genes that enable bacteria survival and spread in hospitals and community. This makes plasmids study a reliable target to control infectious diseases and improve surveillance of antibiotic resistance programs. Current plasmid assembly algorithms and plasmid identification tools are unable to correctly identify all type of plasmids present in a sample, nor their structure.

We introduce PlasmidID, a mapping-based tool that use plasmid database to obtains suitable scaffolds for plasmid identification and reconstruction using assembled contigs coordinates, outputting a summary table with detailed information, a summary image with the most likely plasmids present in a sample and an individual image with full annotation, with emphasis on ARG and incompatibility groups.

A set of simulated whole genome sequencing sequences were generated to test the tool and three high-risk strains responsible for *bla*_{OXA-48}, *bla*_{VIM-1} and *bla*_{KPC-3} dissemination were analyzed, obtaining an average of three plasmid per sample and identifying the ARG-bearing plasmid for each sample.

PlasmidID identifies the most likely plasmids present in a sample, reconstruct them using the bacterial sample sequences provided and draw each one in a self-described image. Since it is database dependent tool, it can identify plasmids of each length and species.

Keywords: Plasmid, Antibiotic resistance, High throughput sequencing

1 Introduction

Plasmids are circular accessory molecules primarily responsible for dissemination of antibiotic resistance genes (ARG), virulence factors and other accessory genes that enable bacteria survival and spread in hospitals. This makes the study of their structure

and composition necessary, since they are a relevant target in treatments for infectious diseases. However, the modular nature of plasmid structure hinders its identification and further comparison using the most usual high-throughput sequencing (HTS) techniques.

HTS is steadily incorporated to clinical microbiology giving more depth and accuracy to diagnosis service, but current available tools based on assembly (de novo assembly and contig alignment) fail its purpose and introduce many false positive results. To solve those issues, we have developed PlasmidID, a mapping-based, assembly-assisted plasmid identification tool which identifies the most likely plasmids present in a sample and use them as scaffold to place the actual sequences in such bacteria.

2 Materials and methods

PlasmidID has been primarily implemented in bash, awk and combining several third-party tools for some steps (bedtools, samtools, prokka and blast). Illumina reads of chromosome and plasmids for *E. coli*, *K. pneumoniae* and *S. enterica* were simulated to adjust PlasmidID default parameters and three different ARG bearing *K. pneumoniae* sequences were tested.

Three databases were fetched, a NCBI reference plasmid database containing 8734 sequences used as scaffold, and an AR (ARG-ANNOT) and replisome (PlasmidFinder) databases for annotation and typing purpose.

3 Results

PlasmidID maps Illumina reads over sequence plasmid database. The most covered sequences are clustered by identity to avoid redundancy and the longest are used as scaffold for plasmid reconstruction. Reads are assembled and the resulting contigs structurally and functionally annotated. All information generated independently from mapping, assembly, annotation and local alignment analyses is gathered and accurately represented in a circular image which allow user to determine plasmidic composition in any bacterial sample analyzed. PlasmidID also generates a summary table which lists all plasmid covered more than 80%, indicating accession number, length, species where the plasmid has been described, description and coverage in every sample analyzed.

Test performance on simulated dataset yielded high accurate results. Plasmid known to belong to each sample were correctly identified (Table 1) and fully reconstructed, including Inc groups and ARG.

A correct plasmid identification was performed on real outbreak samples, furthermore, results suggested the discovery of a not previously described *bla*_{OXA-48}-bearing plasmid (Figure 1).

AC_Number	Length	Species	Description	<i>E. coli</i>	<i>K. pneumo- niae</i>	<i>S. ente- rica</i>
NC_019049.1	4012	<i>E. coli</i>	plasmid pCM959	99.9999		
NZ_CP015856.1	115432	<i>E. coli</i>	strain EDL933-1 plasmid pO157	99.9983		
NZ_CP018246.1	95341	<i>E. coli</i>	strain 472 plasmid pO157	99.9979		
NZ_CP015022.1	95170	<i>E. coli</i>	strain SRCC 1675 plasmid pSRCC 1675	99.9979		
NZ_CP018253.1	95229	<i>E. coli</i>	strain 9000 plasmid pO157	99.9979		
NC_007414.1	92077	<i>E. coli</i>	O157:H7 str. EDL933 plasmid pO157	99.9978		
NZ_CP017670.1	92755	<i>E. coli</i>	strain PA20 plasmid pO157	99.9978		
NZ_CP017435.1	92624	<i>E. coli</i>	O157:H7 strain 1130 plasmid pO157	99.9978		
NZ_CP018244.1	92522	<i>E. coli</i>	strain 350 plasmid pO157	99.9978		
NZ_CP018242.1	92495	<i>E. coli</i>	strain 319 plasmid pO157	99.9978		
NZ_CP018251.1	91789	<i>E. coli</i>	strain 10671 plasmid pO157	99.9978		
NZ_CP018238.1	91420	<i>E. coli</i>	strain 155 plasmid pO157	99.9978		
NZ_DS999999.1	92381	<i>E. coli</i>	O157:H7 str. TW14588 plasmid pTW14588 scf_1117778227960	99.9935		
NZ_CP017252.1	92691	<i>E. coli</i>	strain NADC 5570/86-24/6564 plasmid pO157	99.9935		
NZ_CP018240.1	94170	<i>E. coli</i>	strain 272 plasmid pO157	99.9289		
NZ_CP015021.1	81401	<i>E. coli</i>	strain 28RC1 plasmid p28RC1	99.9177		
NZ_CP006633.1	3319	<i>E. coli</i>	PCN033 plasmid p1PCN033	99.7288		
NZ_AVCD01000003.1	98062	<i>E. coli</i>	O157:H7 str. F8092B plasmid pO157	99.5085		
NZ_CP015847.1	95598	<i>E. coli</i>	O157:H7 strain FRIK2069 plasmid pO157	99.5		
NZ_CP015816.1	95910	<i>E. coli</i>	O157:H7 strain JEONG-1266 plasmid pO157	99.4974		
NZ_CP010305.1	94730	<i>E. coli</i>	O157:H7 str. SS52 plasmid pO157	99.4954		
NC_013010.1	94601	<i>E. coli</i>	O157:H7 str. TW14359 plasmid pO157	99.4884		
NZ_AETX01000217.1	89762	<i>E. coli</i>	O157:H7 str. G5101 plasmid pO157_	86.9533		
NZ_CP010135.1	91607	<i>E. coli</i>	strain D1 plasmid A complete genome	81.3104		
NC_016846.1	111195	<i>K. pneumoniae</i>	subsp. pneumoniae HS11286 plasmid pKPHS2		99.9991	
NC_016839.1	105974	<i>K. pneumoniae</i>	subsp. pneumoniae HS11286 plasmid pKPHS3		99.9981	
NC_016838.1	122799	<i>K. pneumoniae</i>	subsp. pneumoniae HS11286 plasmid pKPHS1		99.9976	
NZ_CP015133.1	26450	<i>K. pneumoniae</i>	strain Kpn555 plasmid pKPN-d6b		99.9962	
NC_016840.1	3751	<i>K. pneumoniae</i>	subsp. pneumoniae HS11286 plasmid pKPHS4		99.9733	
NC_016847.1	3353	<i>K. pneumoniae</i>	subsp. pneumoniae HS11286 plasmid pKPHS5		99.9702	
NC_016841.1	1308	<i>K. pneumoniae</i>	subsp. pneumoniae HS11286 plasmid pKPHS6		99.9235	
NZ_CP011334.1	2954	<i>E. coli</i>	O104:H4 str. C227-11 plasmid		90.5213	
NZ_CP012565.1	112059	<i>K. pneumoniae</i>	strain UCLAOXA232KP plasmid pUCLAOXA232-4		89.0683	
NZ_CP016924.1	111539	<i>K. pneumoniae</i>	isolate 23 plasmid unnamed1		89.0433	
CM003767.1	3652	<i>P. aeruginosa</i>	strain BH6 plasmid pBH6 whole genome shotgun sequence		87.2125	
NZ_CP012427.1	126466	<i>K. pneumoniae</i>	strain KP5 plasmid pSg1-1		86.5837	
NZ_CP008931.1	111693	<i>K. pneumoniae</i>	strain PMK1 plasmid pPMK1-B		85.077	
NZ_CP015755.1	109349	<i>K. pneumoniae</i>	strain W14 plasmid unnamed2		82.6135	
NZ_CP007137.1	64562	<i>E. coli</i>	O145:H28 str. RM12581 plasmid pRM12581		81.661	
NC_003385.1	106516	<i>S. enterica</i>	subsp. enterica serovar Typhi str. CT18 plasmid pHCM2			99.9999
NC_003384.1	218160	<i>S. enterica</i>	subsp. enterica serovar Typhi str. CT18 plasmid pHCM1			99.9977
NZ_LN868946.1	141119	<i>S. enterica</i>	subsp. enterica serovar Senftenberg genome assembly NCTC10384 plasmid: 4			98.852
NC_013365.1	204604	<i>E. coli</i>	O111:H- str. 11128 plasmid pO111_1 DNA			96.7591
NZ_LN868945.1	147787	<i>S. enterica</i>	subsp. enterica serovar Senftenberg genome assembly NCTC10384 plasmid: 3			96.3096
NC_009981.1	208409	<i>S. enterica</i>	subsp. enterica serovar Choleraesuis plasmid pMAK1			95.2819
NC_002305.1	180461	<i>S. typhi</i>	plasmid R27			93.7034
NZ_LN868944.1	727905	<i>S. enterica</i>	subsp. enterica serovar Senftenberg genome assembly NCTC10384 plasmid: 2			92.0865
NZ_CP017724.1	106449	<i>S. enterica</i>	subsp. enterica serovar Stanleyville plasmid pSARB26_01			90.9647
NZ_CP011655.1	109688	<i>C. freundii</i>	strain CAV1741 plasmid pCAV1741-110			89.5403
NC_016825.1	181431	<i>S. enterica</i>	subsp. enterica serovar Typhi str. P-stx-12 plasmid			85.6904
NC_023289.2	239151	<i>E. coli</i>	strain T23 plasmid pEQ1			83.9859

Table 1. Plasmids covered more than 80% of its length for the three simulated samples. Headers include Accession Number, length, species where plasmid has been described, full description and percentage covered. Plasmids are ordered by descending coverage percentage for each sample. Grey cells indicate a plasmid included in the dataset.

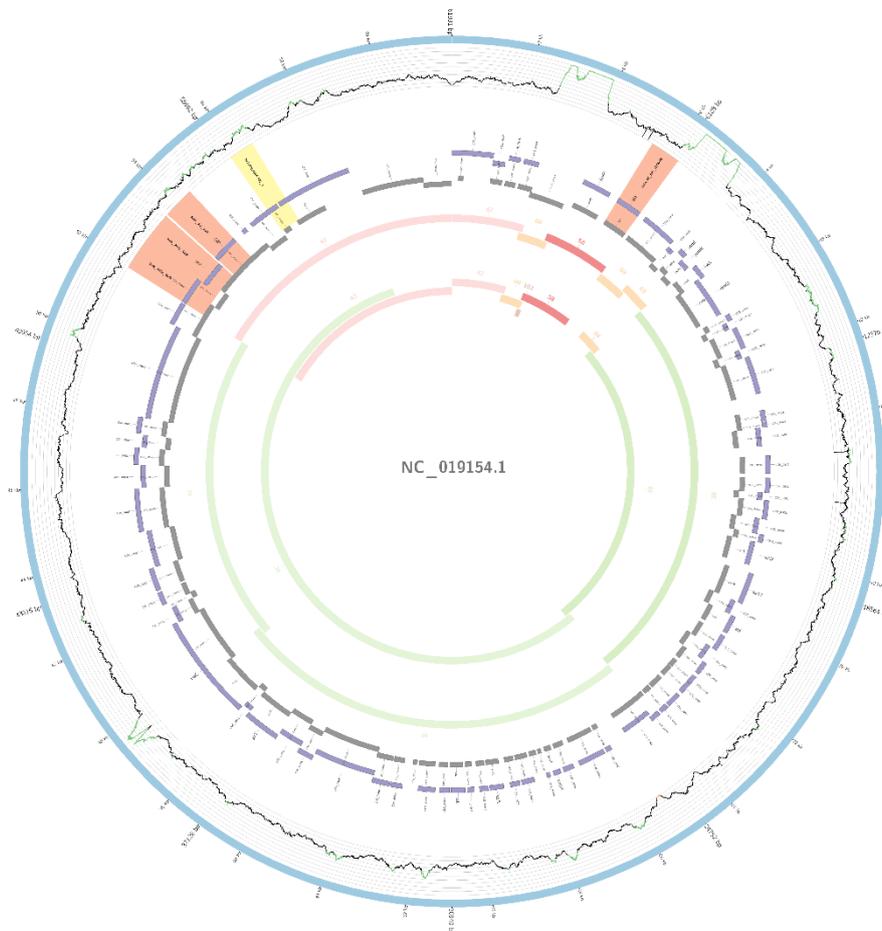


Fig. 1. Plasmid reconstructed over NC_019154 represented as blue circle. Histogram indicates number of reads mapped at each position. Grey and purple box are annotation from reference plasmid and sample contigs respectively. Inner track represents homology region within plasmid and contigs and the full length of those contigs is drawn in the innermost track. Red stripes represent ARG and yellow stripes the incompatibility group.

4 Conclusions

PlasmidID allow plasmids identification in bacterial whole genome sequences, generating images of the most likely plasmids present in a sample, easing interpretation by researchers. Since it is database dependent tool, it works for every plasmid length and species, and can be adapted to other aims such metaplasmid analysis or sequencing technologies as SMRT sequencing.

DB2DB: An enrichment databases approach for medical domain

Zina Nakhla, Kaouther Nouira, and Ahmed Ferchichi

Université de Tunis
BESTMOD Laboratory, Institut Supérieur de Gestion de Tunis

Abstract. The quality of databases are often unsatisfactory, in term of completeness of data. Databases must be up to date to ensure the satisfaction of users. Since there are a large number of databases, an automatic enrichment approach is required. In this paper, we propose a new approach to enrich database called 'DataBase To Database' (DB2DB). This approach aligns the databases and creates an enriched database. DB2DB approach compares the components of two databases, detect the missing components, detect redundancies, combine the constraints, and create a new database. We applied our approach in the medical domain which is characterized by an important volume of scalable information. For experimentation, a platform is developed to test our approach using medical database. As a result we obtain an enrichment database with new components that are either tables, relationship, or records.

Keywords: Enrichment, Database, medical domain.

1 Introduction

Database management systems are used for the storing various kinds of data for different purposes. It enabled information to be efficiently stored and queried. However, databases are not complete, the new elements of knowledge must be continually well-maintained [7]. Data enrichment is a value adding process, where external data from multiple sources is added to the existing data set to enhance the quality and richness of the data. It adds value and improve the quality of your organization's data. Researches in the literature have proposed several approaches to improve database structure and data by merging different data sources [6], using a schema matching [11], using machine learning [1]. The works that have been proposed in literature are limited, it only enriches one component of the database that is usually attributes or records. However, there is a lack of work that concentrates on enriching tables. Also, the approaches in the literature are semi-automatic which necessitates every time the intervention of an expert.

In this context, we propose to use a database for the enrichment of different components of another database called 'DataBase To Database' (DB2DB). Our proposed approach presents a set of enrichment rules which automate the enrichment of database. The proposed enrichment rules treated all the components of the database (tables, records, relationships). It discovers the correspondence

that may exist between elements of two databases. Domain of application used in this paper is medical domain. Our goal is to have an updated database compared to the original one. The new database contains the missing information, and the novelties of the medical domain which are related to the database.

This paper is organized as follows: Section 2 presents works in the literature, which are related to improve databases and methods to enrich databases. Section 3 explains the architecture of our proposed approach. Section 4 describes the enrichment rules. Section 5 presents DB2DB algorithm. Section 6 presents the experiments and the results. Finally, we give conclusion.

2 Related work

Several approaches in literature aimed to enrich database, specially integration data approach and matching schema. Data integration aims to build a unified schema by merging several data sources. Two fundamental steps when performing data integration : The first step consists of identifying similar schema components. The second step consists of transforming data. It will have the same representation once transferred to the global schema, resultant schema must be free from redundancy [6].

Combining matchers approach as hybrid or a composite one, the former combines multiple match criteria simultaneously, the later use several match algorithms individ usually and combine their results [3]. Traditional match techniques, as mentioned above, exploit the schema characteristic and the data value. However additional techniques have been proposed as substitutional choice, since such information are not always interpretable. Kang and Naughton propose Graph-based matching approach that can be useful when schema information is worthless to be used. It consists of exploiting the dependency relationship between schema attributes to construct a dependency graph for each schema to be matched, matching node pairs will be identified using a graph-based matcher [10]. Elmeleegy et al. propose Usage-based approach which detect attributes similarity using database query log. Attributes can be considered as similar if they have similar usage characteristics and occurrence frequencies [8]. Another approach proposed by Nandi and Bernstein used the keyword-query logs to perform schema matching, by considering the results of search queries [2]. Massmann and Rahm proposed an approach called Document content similarity. It creates virtual documents, containing schema instances in order to perform matching by comparing virtual documents using TF/IDF document similarity measure [5]. Gomes de Carvalho et al. proposed Genetic Programming for schema matching. It suggested using a genetic programming in order to identify complex matching. In fact, n:m match is considered to be a challenging issue. However most of the previous studies do not take it into account [4].

In 2016, Anam et al. proposed a new approach named Hybrid Ripple-Down Rules. It address the problem of machine learning algorithms and knowledge engineering approaches. The former cannot deal with the evolutions of schema since classification decisions. The problem of the later arise because rules are

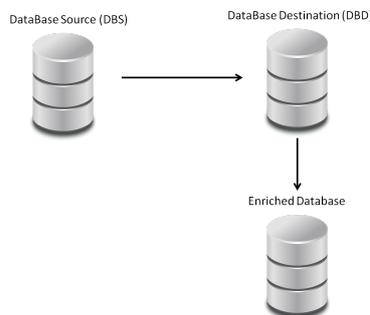


Fig. 1. DB2DB process

constructed by an expert, which is time consuming task and need knowledge acquisition [1]. They propose a hybrid approach that combines machine learning algorithms by using a decision tree and knowledge engineering by adding new rules incrementally when necessary, in order to stop wrong classification and propose a correct one. Recently, Gu et al. proposed Ripple-Down Rules approach. It combines schema matching and record linkage techniques [9]. The results are proved that combining them iterative enhance the match result.

We propose a new approach to automate the enrichment of databases and avoid the limits of approaches presented in literature.

3 DB2DB approach

Usually we found different computer systems related to similar databases. Despite the difference issues of applications, it used nearly similar databases which can contain the same tables, most of the same attributes and records. Each database is created separately, which is a waste of time, effort and storage space. To resolve this issue, simply share the database to optimize the development of application. We can investigate the created databases to build an enriched database (see Fig 1). The main objective of our approach is to automate the enrichment of different components of a database. Our approach focuses on data but also on the database structure. The proposed approach started by calculating the similarity between database source and database destination. Database source is a database used for enrichment, database destination is a database needed to be enriched. We compare the components of databases. Then, we selected the components from the two databases, match the selected components, and unify them creating a new database. The selected components will be compared before matching to avoid redundancy.

4 DB2DB rules

This enrichment of the structure of database concerns the enrichment of table, attribute and relation between tables. If two tables t_{S1} and t_{S2} in DataBase

Source (DBS) are linked. t_D is a table in the DataBase Destination (DBD). If t_{S1} have a similar table in DBD and t_{S2} not have a similar table e.g have the similar name or synonyms and the same attribute. We can add t_{S2} as DBD (see Fig 2). Rule 1 describes this case.

Rule 1:

$$\begin{aligned}
 & \text{When } t_{S1}, t_{S2} \in DBS \text{ and } t_D \in DBD \\
 & \text{and Similarity}(t_{S1}, t_{D1}) \text{ and Relation}(t_{S1}, t_{S2}) \\
 & \text{Then AddTable}(t_{S2}, DBD)
 \end{aligned} \tag{1}$$

Where:

AddTable(T, DB): adds table T to DB.

Relation(T1, T2): returns true if tables T1 and T2 have a relationships between them.

Similarity(T1, T2): returns true if tables T1 and T2 are similar.

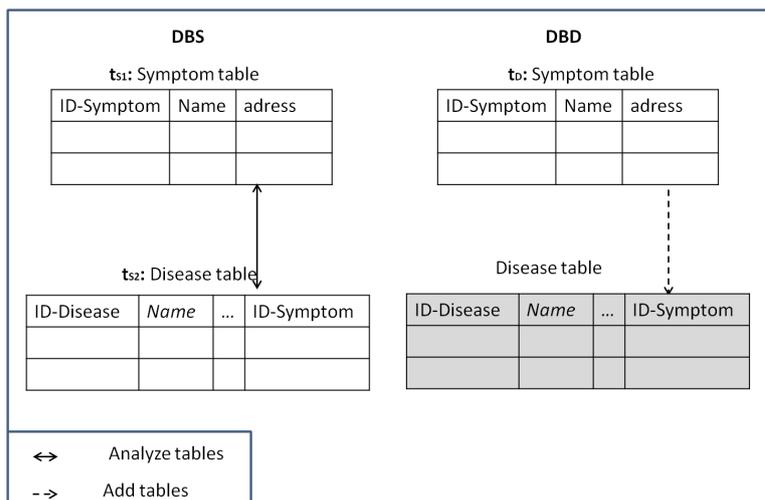


Fig. 2. Rule1: The enrichment of tables.

If two tables in DBS are linked using primary-key and foreign-key and the similar tables appear in the DBD. However, tables in DBD have no link between. Based on the relation in the first database, we can complete the missing relation. We add primary-key in table one and migrate as foreign-key in the second table (see Fig 3). Rule 2 describes this case.

Rule 2:

$$\begin{aligned}
 & \text{When } t_{S1}, t_{S2} \in DBS \text{ and } t_{D1}, t_{D2} \in DBD \\
 & \text{and Similarity}(t_{S1}, t_{D1}) \text{ and Similarity}(t_{S2}, t_{D1}) \\
 & \text{and Relation}(t_{S1}, t_{S2}) \text{ and } \neg \text{RelatedTable}(t_{D1}, t_{D1}) \\
 & \text{Then AddFK}(t_{D1}, t_{D1})
 \end{aligned}
 \tag{2}$$

Where:

Relation(X, Y): returns true if relation exists between two tables.

RelatedTable(T1, T2): returns true if two tables are linked with primary-key and foreign-key.

AddFK(T1, T2): Migates primary key of table T1 as a foreign key in table T2.

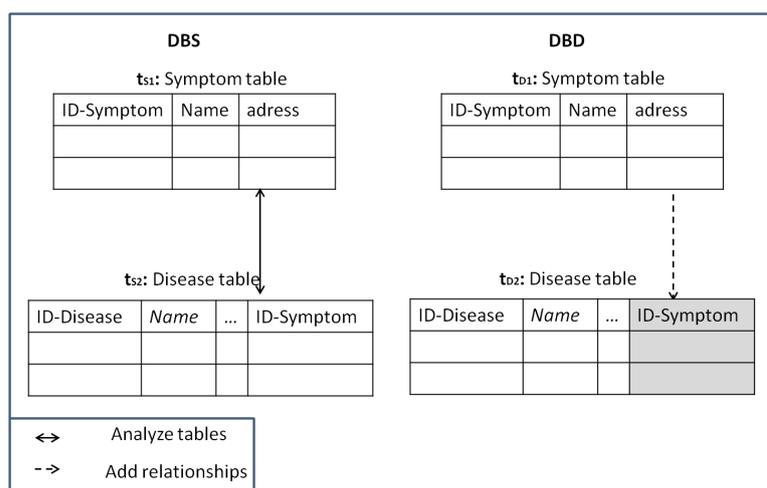


Fig. 3. Rule2: The enrichment of relationships.

We analyze tables and we select the appropriate records for the enrichment of database. The selection of records must follow some conditions: If tables in the databases are similar e.g having the same name or synonyms. The records will be added as a record in the table of DBD (see Fig 4). Also, records of table will be compared to avoid redundancy. Rule 3 describes the enrichment of the database by records.

Rule 3:

$$\begin{aligned}
& \text{When } t_S \in DBS \text{ and } t_D \in DBD \text{ and } \text{Similarity}(t_S, t_D) \\
& \text{and } \text{Redundancy}(\text{record}(t_S), \text{record}(t_D)) \\
& \text{Then } \text{AddRecord}(t_D, \text{record}(t_S)).
\end{aligned} \tag{3}$$

Where:

$\text{Redundancy}(\text{Data1}, \text{Data2})$: compares between two lists of data, and returns data without redundancy.

$\text{AddRecord}(T, R)$: Adds records R in table T.

$\text{Record}(T)$: returns records of table T.

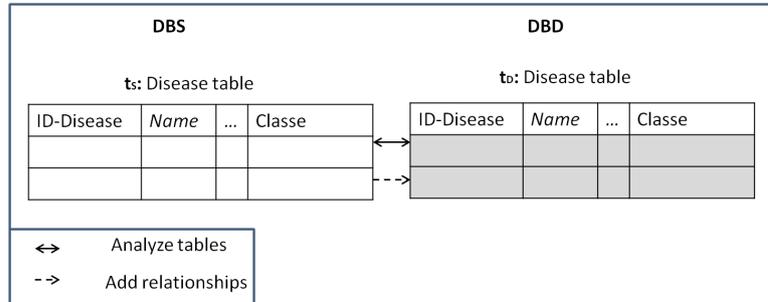


Fig. 4. Rule4: The enrichment of records.

5 DB2DB algorithm

The algorithm is described as follows: Firstly, we browsed all tables of databases and we used rule 1 to add tables in database. Secondly, we used rule 3 to add attributes in table of database. Then, we browsed all relations between databases and used rules 2 to create relation between tables. Finally, we added records of tables in DBS to relative tables in enriched database using rule 4.

6 Experimentation

For the experimentation, we choose medical domain because it is always renewable and it characterized by a large volume of data. We used five different medical databases to test our enrichment rules, and the selection is based on the variation of specialties of each databases. Table 1 shows the number of different components of the selected databases.

The performance of our approach is assessed by the evaluation of generated database: (1) quantitative evaluation and (2) qualitative evaluation.

Algorithm 1 DB2DB

```

1: Input: Database DBS, DBD
2: Output: DataBase Enriched DB
3: Begin
4: for all table  $t$  in DBS and DBD do
5:   Apply Rule1;
6: end for
7: for all relations  $R$  between two tables do
8:   Apply Rule2;
9: end for
10: for all table  $t$  with records set  $r$  do
11:   Apply Rule3;
12: end for
13: Return EnrichedDB

```

Table 1. Medical databases used for the experimentation.

Database	Table	Record	Attribute
DB1: Disease Database	10	43	32
DB2: Allergy Database	15	53	63
DB3: Cardiovascular Database	11	21	43
DB4: Surgical Database	9	28	30
DB5: Laboratory analysis Database	12	27	49

6.1 Quantitative evaluation

Calculate the similarity We started our evaluation by calculating the similarity between databases. To measure the semantic similarity of databases, we browse the components of databases and, determine the similarity of components based on the semantic similarities using the SNOMED-CT ontology¹ to subtract similarity between concepts, which is the most comprehensive and precise clinical health terminology product in the world. Similarity value $Sim(t)$ calculates the value of similarity of term t related to databases, $0 \leq Sim(t) \leq 1$. t can be a name of any database components (tables, attributes, records).

$$\begin{cases} Sim(t) = 1, & \text{if } t \text{ has a similar components.} \\ Sim(t) = 0, & \text{if } t \text{ has not a similar components.} \\ Sim(t) = 0.5, & \text{if } t \text{ has a synonym.} \\ Sim(t) = 0.75, & \text{if } t \text{ has a other form.} \end{cases} \quad (4)$$

¹ <http://www.snomed.org/snomed-ct>

$$Similarity(A, B) = \frac{\sum_{t \in T_A \cap T_B}^N Sim(t)}{N} \quad (5)$$

Where

N : Number of similar term to t .

If two databases are not similar or the percentage of similarity is inferior of 50%, we cannot obtain an enriched databases. To enrich a DBD, we need a DBS in the same domain an with a high degree of similarity with DBD. Table2 presents databases, we choose DB1 and DB2 as database source of enrichment because it has greater number of components than others.

Table 2. The rate of similarity between tables.

Enrichment code	Database source	Database destination	rate of similarity (%)
Enrich#1	DB1	DB3	76
Enrich#2	DB1	DB4	60
Enrich#3	DB2	DB3	80
Enrich#4	DB2	DB4	78

Calculate the enrichment rate We evaluate the quantity of the new information stored in the database. We have computed the enrichment rate of the database, which is the percentage of changement of the number of database components after enrichment by comparison to number of components before the enrichment. The sum of component consists on tables, records, attribute, relationships between tables using primary and foreign key. E_2 computed the sum of components of database destination after the enrichment. E_1 computed the sum of changements in database destination before the applying of the enrichment rules.

$$(E_1 - E_2)/E_2 * 100 \quad (6)$$

Figure 5 illustrates the enrichment rates for each databases combination. Table 3 resumes the enriched results obtained.

According to the sketched histograms, we can point out that most of enrichment rates of databases using DB_1 are height then DB_2 . We remark that $Enrich_3$ has the highest rate of enrichment for DBS and DBD, which explained by the height rate of similarity (80%) between the components of the databases. Besides DB_4 have a small number of tables which prove that needs enrichment. In contrast, $Enrich_2$ have a small rater of changement when we use DB_2 . The

explanation is that DB_4 and DB_1 both contained less of similar components and does not require a high rate of change.

Table 3. The enrichment rates for databases using DB1 and DB2.

Enrichment code	E_1	E_2	(%)
Enrich#1	145	167	13.17
Enrich#2	98	105	6.66
Enrich#3	84	116	27.5
Enrich#4	75	88	14.77

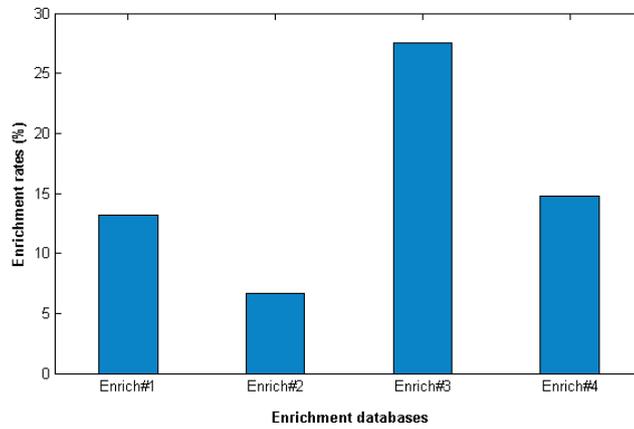


Fig. 5. Enrichment rates.

6.2 Qualitative evaluation

The qualitative evaluation is based on the satisfaction level of the users of database before and after enrichment. We select four criteria of satisfaction which are (1) Accessibility: the ability to access the database, (2) Accuracy: the degree to which the data mirrors the characteristics of the real world object or objects it represents, (3) Completeness: the absence of blank and missed values, and (4) Consistency: the absence of difference, when comparing two or more representations of a thing against a definition. We asked twenty experts to judge the information quality of new databases compared to the last version of database. To judge we have two dimensions of satisfaction (high and low). Table 4 presents the number of expert for each dimension and for each criteria. Based on table 4, we conclude that the number of experts, which are satisfied

with resulted databases, is higher than the number of experts which are not satisfied of the quality of databases. It proves that users have not problems in the access to databases. Also, it proves that data in databases are correct, accurate, complete, and characterized by the absence of ambiguity.

Table 4. Level of satisfaction of the experts according to quality criteria for each database.

Information quality criteria	Enrich#1		Enrich#2		Enrich#3		Enrich#4	
	High	Low	High	Low	High	Low	High	Low
Accessibility	20	00	20	00	19	01	20	00
Accuracy	13	07	15	05	13	07	10	10
Completeness	20	00	18	02	16	04	19	01
Consistency	16	04	17	03	17	03	18	02
Total	59	14	70	16	65	15	67	13

Discussion

We can summarize the performance of our approach based on above evaluations of experiments. Both quantitative and qualitative evaluations highlights the efficiency of our automatic approach of database enrichment. Table 3 presents the results of quantitative evaluation which are the rate of enrichment for each database combination. The obtained results show the performance of our approach for different databases. Then, we evaluate the quality of database after the enrichment. Table 4 presents the level of satisfaction of experts based on four quality criteria. This table shows that experts are satisfying with the enriched database in the most of criteria. Through all the evaluation of experiments, we demonstrate that our approach has enhanced the enrichment of database significantly.

7 Conclusion

We proposed an approach to enrich the database using other database. The goal was to achieve a way to analyze and compare between the components of two databases. We presented in this paper a set of enrichment rules which enrich records of database and the structure of database. We developed a platform to apply rules and we use five medical databases. We evaluate the resulted databases with quantitative and qualitative evaluations. The quantitative evaluation shows that the number of database component grows after the application of rules. The qualitative evaluation shows that resulted database meets the quality criteria. As a future work, we will test our approach using voluminous databases. And we will deal with the addition of new semantic relations using the inference rules.

References

1. S. Anam, Y. S. Kim, B. H. Kang, and Q. Liu. Adapting a knowledge-based schema matching system for ontology mapping. In *Proceedings of the Australasian Computer Science Week Multiconference, ACSW '16*, pages 27:1–27:10, New York, NY, USA, 2016. ACM.
2. P. A. Bernstein, M. Dashti, T. Kiefer, and D. Maier. Indexing in an actor-oriented database. In *CIDR 2017, 8th Biennial Conference on Innovative Data Systems Research, Chaminade, CA, USA, January 8-11, 2017, Online Proceedings*, 2017.
3. B. Chien and S. He. A hybrid approach for automatic schema matching. In *International Conference on Machine Learning and Cybernetics, ICMLC 2010, Qingdao, China, July 11-14, 2010, Proceedings*, pages 2881–2886, 2010.
4. M. G. de Carvalho, A. H. F. Laender, M. A. Gonçalves, and A. S. da Silva. An evolutionary approach to complex schema matching. *Inf. Syst.*, 38(3):302–316, 2013.
5. H.-H. Do and E. Rahm. Coma: A system for flexible combination of schema matching approaches. In *Proceedings of the 28th International Conference on Very Large Data Bases, VLDB '02*, pages 610–621. VLDB Endowment, 2002.
6. A. Doan and A. Y. Halevy. Semantic integration research in the database community: A brief survey. *AI Magazine*, 26(1):83–94, 2005.
7. E. M. Dobrescu and D. M. Pociovliteanu. Some observations regarding the international databases. *Procedia Economics and Finance*, 32:841 – 849, 2015.
8. H. Elmeleegy, M. Ouzzani, and A. K. Elmagarmid. Usage-based schema matching. In *Proceedings of the 24th International Conference on Data Engineering, ICDE 2008, April 7-12, 2008, Cancún, México*, pages 20–29, 2008.
9. B. Gu, Z. Li, X. Zhang, A. Liu, G. Liu, K. Zheng, L. Zhao, and X. Zhou. The interaction between schema matching and record matching in data integration. *IEEE Trans. Knowl. Data Eng.*, 29(1):186–199, 2017.
10. J. Kang and J. F. Naughton. On schema matching with opaque column names and data values. In *Proc. 2003 ACM SIGMOD Int. Conf. Manag. data - SIGMOD '03*, page 205, New York, New York, USA, June 2003. ACM Press.
11. E. Rahm and P. A. Bernstein. A survey of approaches to automatic schema matching. *VLDB JOURNAL*, 10:2001, 2001.

THE IMPORTANCE OF USING QUALITY CONTROL SAMPLES TO APPLY DATA PROCESSING METHODS IN LARGE FINGERPRINTING METABOLOMIC STUDIES. A PRACTICAL CASE WITH 288 SAMPLES OF SYSTEMIC AUTOIMMUNE DISEASES.

FERNÁNDEZ-OCHOA, ÁLVARO^{1,2} (0000-0001-5547-3209), BORRÁS-LINARES; ISABEL² (0000-0002-5227-9002); QUIRANTÉS-PINÉ, ROSA² (0000-0002-2107-6333); PRECISESADS Clinical Consortium; ALARCÓN RIQUELME, MARTA E³ (0000-0002-7632-4154); SEGURA-CARRETERO, ANTONIO^{1,2} (0000-0002-5564-5338).

¹ Department of Analytical Chemistry, Faculty of Sciences, University of Granada, Spain; alvaroferochoa@ugr.es

² Research and Development of Functional Food (CIDAF), Health Science Technological Park, Granada, Spain

³ Centre for Genomics and Oncological Research (GENYO) Pfizer – University of Granada – Andalusian Government, Health Science Technological Park, Granada, Spain

Metabolomics is considered the final step of the ‘omics’ cascade whose aim is to study all low molecular weight molecules present in biological systems. The description of metabolome provides a portrait of the metabolic state of individuals at a certain point in time and the analysis of the metabolic profile may give insight into the biochemical consequences of disease. In this way, this tool allows to find alterations and interactions in the organism due to different conditions or causes [1]. Most of studies carried out in this field have focused on the study of human diseases in order to know the involved biochemical pathways and to find biomarkers that allow the improvement in the diagnosis, prognosis and treatments [2, 3].

Among the different stages of the metabolomic workflow, the data processing is one of the most important. The long duration of the analytical sequences makes that the provided data have to be well processed in order to obtain statistic meaningful and valid and robust biological interpretations.

In this work, we present the utility of using quality control samples (QCs) for data processing in large fingerprinting metabolomic studies. This methodology was applied in a real study, which was a part of the PRECISESADS project (www.precisesads.eu), in order to explore different systemic autoimmune diseases. These pathologies were Systemic Lupus Erythematosus (SLE), Systemic Sclerosis (SSc), Sjögren’s Syndrome (SjS), Mixed Connective Tissue Disease (MCTD), Undifferentiated Connective Tissue Disease (UCTD), Rheumatoid Arthritis (RA) and Antiphospholipid Syndrome (PAPS). In this study, a total of 288 plasma and 288 urine samples

were analysed by high performance liquid chromatography coupled to mass spectrometry (HPLC-ESI-QTOF-MS).

QCs of plasma and urine were prepared by mixing equal volumes (20 μ l) from each case sample. Those QCs were analysed throughout the analytical sequence every 5 case samples.

Firstly, this sample was used in order to check the analytical reproducibility by unsupervised Principal Component Analysis (PCA) according to the distribution of the QCs. The PCA scores allowed to detect significant drifts due to the different batches of analysis (between-batch effect) and the injection order (within-batch effect). Thus, normalization strategies have to be applied in order to suppress both effects. Therefore, a normalization strategy was applied using the batch effect correction tool in Metaboanalyst software [4] based on empirical Bayes method [5]. In addition, normalization procedures based on the sum of areas of the QCs were also applied to correct the injection order effect. After applying these normalization procedures, PCA analysis were remade showing a good grouping of the QCs after processing. Other utility of the QCs was applied in the Peak Peaking step. Automatic peak finding and alignment were carried out with *Agilent MassHunter Profinder B.06.00* software. In this step, QCs were used as a representative sample in which all endogenous metabolites should be present. Due to the large number of sample files and their size, the chosen strategy was to perform the molecular feature extraction in the QCs files acquired along the different analytical batches. The molecular features found in QCs were used to find them by a targeted way in the tested samples.

QCs were also used in the filtering steps. The molecular features with high variability in QCs (relative standard deviation higher than 30%) or with a percentage of missing values higher than 25 %, were removed. An additional application of the QCs was to perform a tandem mass spectrometry (MS/MS) analysis in order to obtain fragmentation patterns of the metabolites. This information was necessary in order to unambiguously identify the significant features found for the different diseases comparing with data from different online databases.

References:

1. Agin, A., Heintz, D., Ruhland, E., Chao de la Barca, J.M., Zumsteg, J., Moal, V., Gauchez, A.S., Namer, I.J.: Metabolomics - an overview. From basic principles to potential biomarkers (part 1), <http://linkinghub.elsevier.com/retrieve/pii/S0928125815003964>, (2016)
2. Johnson, C.H., Ivanisevic, J., Siuzdak, G.: Metabolomics: Beyond biomarkers and towards mechanisms. *Nat. Rev. Mol. Cell Biol.* 17, (2016). doi:10.1038/nrm.2016.25

3. Wang, X., Chen, S., Jia, W.: Metabolomics in Cancer Biomarker Research. *Curr. Pharmacol. Reports.* 2, (2016). doi:10.1007/s40495-016-0074-x
4. Xia, J., Wishart, D.S.: Using MetaboAnalyst 3.0 for Comprehensive Metabolomics Data Analysis. *Curr. Protoc. Bioinformatics.* 55, 14.10.1-14.10.91 (2016). doi:10.1002/cpbi.11
5. Johnson, W.E., Li, C., Rabinovic, A.: Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics.* 8, 118–127 (2007). doi:10.1093/biostatistics/kxj037
6. Chetwynd, A.J., Abdul-Sada, A., Holt, S.G., Hill, E.M.: Use of a pre-analysis osmolality normalisation method to correct for variable urine concentrations and for improved metabolomic analyses. *J. Chromatogr. A.* 1431, 103–110 (2016). doi:10.1016/j.chroma.2015.12.056

Acknowledgments:

This work has received support from the Innovative Medicine Initiative joint undertaking under grant agreement n° 115565, resources of which are composed of the financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and the EFPIA companies' in kind contribution.

Annex. Members from the PRECISESADS Clinical Consortium.

Clinical center	Principal investigator	Clinicians
Hospital Regional de Málaga, Servicio Andaluz de Salud , Málaga (Spain)	Enrique de Ramón Garrido	
Hospital Universitario San Cecilio, Servicio Andaluz de Salud, Granada (Spain)	Norberto Ortego; Enrique Raya	María Concepción Fernández Roldán; José Luis Callejas Rubio; Raquel Ríos Fernández; Inmaculada Jiménez Moleón;
Hospital Universitario Reina Sofia Andaluz de Salud, Córdoba (Spain)	Eduardo Collantes	Rafaela Ortega-Castro M ^ª Angeles Aguirre-Zamorano Alejandro Escudero- Contreras M ^ª Carmen Castro-Villegas
Centre Hospitalier Universitaire de Brest, Hospital de la Cavale Blanche, Brest, (France)	Jacques-Olivier Pers	Alain Saraux Valérie Devauchelle-Pensec Divi Cornec Sandrine Jousse-Joulin
Fondazione IRCCS Ca Granda Ospedale Maggiore Policlinico, Milano (Italy)	Lorenzo Beretta	
Deutsches Rheuma-Forschungszentrum Berlin, (Germany)	Falk Hiepe	
Hospitaux Universitaires de Genève (Switzerland)	Carlo Chizzolini	
Centro Hospitalar do Porto (Portugal)	Carlos Vasconcelos	Ana Campar António Marinho Fátima Farinha Isabel Almeida Mariana Brandão Raquel Faria
Medizinische Hochschule Hannover (Germany)	Torsten Witte	Niklas Baerlecken
Katholieke Universiteit Leuven (Belgium)	Rik Lories	Ellen De Langhe
Université catholique de Louvain (Belgium)	Bernard Lauwerys	
Università degli studi di Milano, (Italy)	Pier Luigi Meroni	
Klinikum Der Universitaet Zu Koeln, (Germany)	Nicolas Hunzelmann	Doreen Belz
Medizinische Universität Wien, (Austria)	Georg Stummvoll	Michael Zauner, Michaela Lehner
University of Szeged, (Hungary)	Laszló Kovács	Attila Balog, Magdolna Deák, Márta Bocskai, Sonja Dulic, Gabriella Kádár
Hospital Clinic I Provincia, Institut d'Investigacions Biomèdiques August Pi i Sunyer, Barcelona (Spain)	Ricard Cervera	Ignasi Rodríguez-Pintó, Gerard Espinosa
Hospital Universitario Marqués de Valdecilla, Servicio Cántabro de Salud, Santander (Spain)	Miguel A. González-Gay	Ricardo Blanco Alonso Alfonso Corrales Martínez
Andalusian Public Health System Biobank	Blanca Miranda	Rocío Aguilar Quesada
Project Office - Recruitment and data follow up	Jacqueline Marovac & Tania Gomes Anjos	

Detecting Cancer-Associated Epistatic Gene Variants in Lung Adenocarcinoma

Jaume Sastre Tomas 1,2, Jairo Rocha Cárdenas 1,2, Damiá Heine-Suner 2, and Emidio Capriotti 3

1. Department of Mathematics and Informatics, Universidad de las Islas Baleares (UIB), Palma, Spain

2. Instituto de Investigacion Sanitaria de las Islas Baleares (Idisba), Palma, Spain

3. Department of Pharmacy and Biotechnology (FaBiT), University of Bologna (Unibo), Bologna, Italy

{jaume.sastre, jairo.rocha}@uib.es

{emidio.capriotti}@unibo.edu

{damia.heine}@ibsallut.es

<http://www.uib.es/es/recerca/estructures/grups/grup/BIOCOM/>

Keywords: cancer genomics, driver mutations, epistatis, gene-gene interactions

Abstract. Cancer is a complex disease driven by several factors, ranging from point mutations that are accumulated during lifetime (somatic), inherited variants (germline), large rearrangements such as deletions, insertions or translocations (structural variants) or epigenetic factors [1, 2].

Somatic mutations in cancer are very heterogeneous, very few mutations are shared by patients with the same cancer, even those with the same subtype of cancer. Sets of genes that are commonly involved in the same cancer pathway tend not to be mutated in the same patient [3].

We have analyzed paired mutational events in lung adenocarcinoma to detect pairs of genes frequently mutated in tumor samples potentially associated to the insurgence and progression of the disease.

We collected a dataset of samples of Lung Adenocarcinoma released by TCGA (The Cancer Genome Atlas) consortium [4]. We have used Varscan2 to perform the variant calling analysis starting from the original BAM files [5]. The version hg38 of the human genome was used as a reference [6]. After a filtering procedure for removing duplicated sequencing data from same patient, we obtained a set of 405 unique paired samples (normal/tumor). The annotation of mutations was performed using Annovar [7]. We have collected the allele frequency (AF) data from gnomAD, 1000 Genomes, ExAC databases and we associated to each variant the maximum AF value. If no AF is reported in any of the three databases its value was set to 0.

In our analysis we assumed that all rare variants affecting protein sequence (VAPs) have pathogenic effects. Thus, for each sample, we col-

lected the VAPs (nonsynonymous Single Nucleotide Variants, frameshift-deletion, frameshift-insertion, stop-gain, stop-loss, nonframeshift-deletion and nonframeshift-insertion) with allele frequency $\leq 0.5\%$. To filter-out possible sequencing errors and artifacts we considered only variants with minimum coverage of 10 and supported by at least 5% of the reads for the alternative allele.

In our work we analyzed co-occurring mutational events affecting the function of pairs of genes. We assumed that a protein coding gene loses its function when a VAP is present. In the analysis of the LUAD dataset we found a total number ~ 17.300 genes with at least one VAP across all samples. To reduce the number of possible gene pairs to be tested, we prioritized cancer-associated genes using a modified version ContrastRank algorithm [8]. This procedure allowed to select a set of ~ 1.400 significantly mutated genes ($p\text{-value} \leq 0.05$) in case (tumor samples) vs control (normal samples).

Starting from this set of genes, we tested all possible gene pairs performing two Fishers Exact Tests (FETs). The first test compares the fraction of patients with paired VAPs in normal vs tumor samples (see Fig. 1 Panel A). The second test calculates a contingency matrix for each gene pair reporting the number tumor samples with the four possible mutational states obtained by the combination of the two genes (see Fig. 1 Panel B).

A		$\overline{\text{EGV}}$	EGV
Normal	$N_{\overline{\text{EGV}}}^{\text{normal}}$	$N_{\text{EGV}}^{\text{normal}}$	
Tumor	$N_{\overline{\text{EGV}}}^{\text{tumor}}$	$N_{\text{EGV}}^{\text{tumor}}$	

B		$\overline{\text{G1}}$	G1
	$\overline{\text{G2}}$	$N_{\overline{\text{G1,G2}}}^{\text{tumor}}$	$N_{\text{G1,G2}}^{\text{tumor}}$
	G2	$N_{\overline{\text{G1,G2}}}^{\text{tumor}}$	$N_{\text{EGV}}^{\text{tumor}}$

Fig. 1. Contingency matrices calculated for evaluating significant Epistatic Gene Variants (EGVs). Panel A: Contingency matrix for the comparison of the frequency of EGVs in normal and tumor samples. Panel B: Contingency matrix for the comparison of EGV states in tumor samples. G1: Gene 1. G2: Gene 2. N: number of samples with or without variants affecting protein sequence (VAP). $N_{\text{EGV}} = N_{\text{G1,G2}}$.

Using FETs we selected a set of ~ 27.000 Epistatic Gene Variants (EGVs) with both $p\text{-value} \leq 0.05$. In a second step, we identified significant changes in the survival rates between patients with and without a specific epistatic gene interaction. Calculating the log-rank statistics from Kaplan-Meier curves [9] for the groups patients with and without the selected epistatic gene interactions, we found 441 EGVs with signif-

icant decrease in the survival rate ($p\text{-value} \leq 0.05$). These EGVs were formed by 267 unique genes.

Mapping our final set of 441 EGVs on the human protein-protein interaction network from iRefIndex [10] we selected a set of 140 gene-gene interactions that contained the aforementioned pairs of genes in either the source or the target of the interaction with a maximum one gene in the shortest path. In Figure 2 we draw the largest module of our network (Panel A). The main hubs in our network are TP53 and EGFR with 195 and 188 interactions respectively (Fig. 2 Panel B). These are two well-known genes that act as tumor suppressor and oncogene respectively in many cancer types. In our network, among other hub genes, we found other two cancer-associated genes such as KEAP1 and STK11 with 13 and 6 interactions respectively.

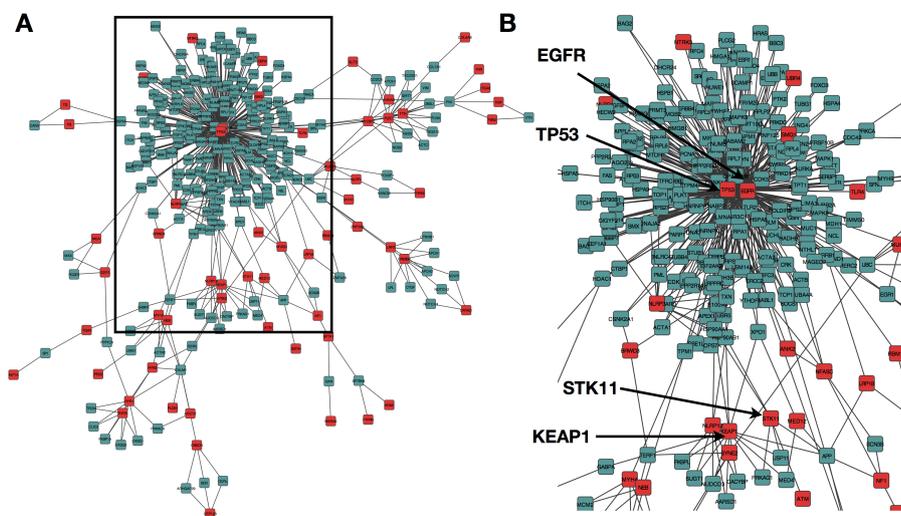


Fig. 2. Main module of our protein-protein interaction network obtained matching the set of 441 Epistatic Gene Variants with the protein-protein interaction from iRefIndex database (Panel A). Panel B zoom on the main hubs of the network.

With our analysis we detected a set of cancer-associated Epistatic Gene Variants with significant decrease of the survival rate in Lung Adenocarcinoma. These Epistatic Gene Variants map on a subnetwork of protein-protein interaction including two well-known hubs (TP53 and EGFR) for cancer-associated gene network. We anticipate that our findings will be used for predicting the survival rate of a patient from the analysis of its mutation landscape.

References

1. B. Vogelstein, N. Papadopoulos, V. E. Velculescu, S. Zhou, L. A. D. Jr., and K. W. Kinzler, "Cancer genome landscapes," *Science*, pp. 1546–1558, March 2013.
2. L. B. Alexandrov, S. Nik-Zainal, D. C. Wedge, S. A. J. R. Aparicio, S. Behjati, A. V. Biankin, G. R. Bignell, N. Bolli, A. Borg, A.-L. Borresen-Dale, S. Boyault, B. Burkhardt, A. P. Butler, C. Caldas, H. R. Davies, C. Desmedt, R. Eils, J. E. Eyfjord, J. A. Foekens, M. Greaves, F. Hosoda, B. Hutter, T. Ilcic, S. Imbeaud, M. Imielinski, N. Jager, D. T. W. Jones, D. Jones, S. Knappskog, M. Kool, S. R. Lakhani, C. Lopez-Otin, S. Martin, N. C. Munshi, H. Nakamura, P. A. Northcott, M. Pajic, E. Papaemmanuil, A. Paradiso, J. V. Pearson, X. S. Puente, K. Raine, M. Ramakrishna, A. L. Richardson, J. Richter, P. Rosenstiel, M. Schlesner, T. N. Schumacher, P. N. Span, J. W. Teague, Y. Totoki, A. N. J. Tutt, R. Valdes-Mas, M. M. van Buuren, L. van 't Veer, A. Vincent-Salomon, N. Waddell, L. R. Yates, A. P. C. G. Initiative, I. B. C. Consortium, I. M.-S. Consortium, I. PedBrain, J. Zucman-Rossi, P. Andrew Futreal, U. McDermott, P. Lichter, M. Meyerson, S. M. Grimmond, R. Siebert, E. Campo, T. Shibata, S. M. Pfister, P. J. Campbell, and M. R. Stratton, "Signatures of mutational processes in human cancer," *Nature*, vol. 500, pp. 415–421, Aug. 2013.
3. B. J. Raphael, "Structural variation and medical genomics," *PLoS Computational Biology*, vol. 8, December 2012.
4. NCI and the NHGRI, "The cancer genome atlas homepage," 2009.
5. D. C. Koboldt, Q. Zhang, D. E. Larson, D. Shen, M. D. McLellan, L. Lin, C. A. Miller, E. R. Mardis, L. Ding, and R. K. Wilson, "VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing," *Genome Research*, vol. 22, pp. 568–576, Jan. 2012.
6. T. . G. P. Consortium, "A global reference for human genetic variation," *Nature*, vol. 526, p. 68, Sept. 2015.
7. K. Wang, M. Li, and H. Hakonarson, "AnnoVar: functional annotation of genetic variants from high-throughput sequencing data," *Nucleic Acids Research*, vol. 38, pp. e164–e164, June 2010.
8. R. Tian, M. K. Basu, and E. Capriotti, "Contrastrank: a new method for ranking putative cancer driver genes and classification of tumor samples," *Bioinformatics*, vol. 30, pp. 572–578, 2014.
9. K. J. Jager, P. C. van Dijk, C. Zoccali, and F. W. Dekker, "The analysis of survival data: the kaplan-meier method," *Kidney International*, vol. 74, no. 5, pp. 560–565, 2013.
10. S. Razick, G. Magklaras, and I. M. Donaldson, "irefindex: A consolidated protein interaction database with provenance," *BMC Bioinformatics*, vol. 9, p. 405, Sept. 2008.

GPU-Low-Energy Tracking of the Left Ventricle in the Cloud

Sidi Ahmed Mahmoudi¹, Mohammed Ammar², Mohammed Amin Belarbi¹,
Amine Abbou², and Saïd Mahmoudi¹

¹ Faculty of Engineering. University of Mons. 9, rue de Houdain. Belgium
sidi.mahmoudi@umons.ac.be

² University of boumerdes, faculty of engineering sciences; department of engineering
System, Algeria

Abstract. Left ventricle tracking in ultrasound images and videos presents a very important application for cardiac diagnostic. This process is difficult due to two main reasons: 1. The inherent problems of ultrasound videos (i.e. low contrast, signal dropout, etc.). 2. The complex process of configuration and exploitation of the related solutions when exploiting high performance software and hardware. In this paper, we propose real-time method that offer an accurate left ventricle segmentation and tracking, based on optical flow estimation, convex hull and spline interpolation algorithms. We propose also GPU-low-energy solution that can exploit the high computing power of embedded GPUs. This allowed to provide a real-time processing of videos within a prototype (Tegra mobile GPU) that consumes low energy compared to normal GPUs. Moreover, our solution is adapted for exploiting cloud platforms in order to offer to users an access to the different algorithms without the need to download, install and configure software or hardware. Our experiments are conducted using a set of 17 normal and 31 disease hearts ultrasound video sequences. The related results achieved automatic and real-time left ventricle detection and tracking with a rate of 93 % of success

Key words : left ventricle detection and tracking, ultrasound videos, optical flow, Embedded GPU, CUDA, Cloud computing.

1 Introduction

Cardiovascular diseases present one of the main leading causes of death in the world. Therefore, cardiac diagnosis have got increasing importance for cardiologists. In literature, there are many modalities used for exploring and analyzing cardiac structures such as: magnetic resonance (MRI), X-ray, ultrasound, or computed tomography (CT) images. Actually, the ultrasound imaging are the most used in clinical practice since they offer real time visualization of the heart motion. Several methods have been proposed to delineate left ventricle within ultrasound imaging. The latter are composed of several image processing algorithms that make the calculation time so elevated. As consequence, the constraint of real time processing cannot be satisfied when using high definition

videos. In this context, graphic processing units present an efficient solution, which is seriously hampered by the high costs of data transfer between CPU and GPU memories. Therefore, we developed a version that exploits in an efficient way, all the available computing units within embedded graphic cards "NVIDIA TEGRA TK1"³. The latter offer the possibility to provide a fast, low-energy and portable solution that does not require the use of a computer. This solution is also adapted for exploiting CPU or GPU within cloud platforms in order to facilitate its use even if users do not dispose of mobile GPUs. The remainder of the paper is organized as follows: Section 2 presents the related works, while the third Section is devoted to describe the proposed approach for left ventricle segmentation and tracking using mobile GPUs and cloud platforms. Finally, conclusion and future works are presented in the last Section.

2 Related work

Left ventricle segmentation and tracking in echocardiographic sequences presents a very important task in cardiac motion abnormalities detection. Many algorithms have been proposed to track the left ventricle such as presented in [1] [2], where authors developed an automatic recognition algorithm for detecting the contour of left ventricular cavity. The method is based on considering all of the combinations between motion vectors of the ventricular wall in order to extract intersection points. This software was applied on 40 normal and 113 disease hearts where automatic calculation of ejection fraction was agreed with the results of conventional method by a sonographer. Tavakoli et al [3] proposed an evaluation of different optical flow methods. Nine different algorithms namely Lucas-Kanade, Horn-Schunck, Fleet-Jepson, Anandan, Nagel, Uras et al., Black et al, correlation and MAD block matching are implemented and applied on simulated, synthetic (rotating phantom) and real ultrasonic images. Authors found that the best results are achieved using Black et al. technique. The angular and amplitude errors obtained are 8 degrees per frame and 8.2%, respectively. Papademetris et al [4], proposed a methodology for estimating cardiac deformations from 3D echocardiography (3DE). A dense motion field is used to calculate the deformation of the heart wall in terms of strain in cardiac specific directions. The strains obtained using this approach in open-chest dogs before and after coronary occlusion, show good agreement with previously published results in the literature. They also exhibit a high correlation with strains produced in the same animals using implanted sonomicrometers. This proposed method provides quantitative regional 3D estimates of heart deformation from ultrasound images. Dietenbeck et al, choose to use a extended level set method [5] in order to track the whole myocardium in echocardiographic sequences. The accuracy and robustness of the proposed method is evaluated by comparing the obtained-segmentation with experts references and to another state-of-the-art method on a dataset of 15 sequences (900 images) acquired in three echocardiographic views. They show that the algorithm provides results that are consistent with

³ <http://www.nvidia.com/object/jetson-tk1-embedded-dev-kit.html>

the inter-observer variability and outperforms the state-of-the-art method. Recently, we developed a Multi-GPU implementation [6] [7] of sparse optical flow computation that allowed real time motion tracking with high definition videos. The latter is exploited within our method of left ventricle tracking.

On the other hand, recent development of computer vision platforms have been significantly influenced by the emergence of a growing number and accessible cloud computing platforms hosted by large-scale IT-companies (AWS, GCP, Azure). They enabled the development of a variety of cloud interfaces, which makes abstraction on the complexity behind computer vision application. The latter use a specific workflow for cloud architectures [8] which gives access to a high computing power without the need of a low-level software programming or any hardware adaptation. CloudCV [9] is an example of a cloud-based and distributed computer vision platform composed of three parts; an AI-as-a-service platform that enables researchers to easily convert their deep learning models to web service and call them by a simple API, a drag and drop collaborative platform for building models, an evaluation server for comparing different AI and computer vision algorithms (for example for challenges). Other cloud commercial applications such as Amazon Rekognition, Watson Visual Recognition (by IBM) or Clarifai, deliver APIs, which focus on specific computer vision tasks for image and video understanding.

In this paper, we developed a version that exploits in an efficient way, all the available computing units within embedded graphic cards "NVIDIA TEGRA TK1". The latter offers the possibility to provide a fast, low-energy and portable solution that does not require the use of a computer. This solution is also adapted for exploiting CPU or GPU within cloud platforms in order to facilitate its use even if users do not dispose of mobile GPUs.

Our main contributions can be summarized in two points:

1. Low-energy and fast implementation of left ventricle detection and tracking as a result of exploiting embedded graphic cards "NVIDIA TEGRA TK1" in parallel.
2. Cloud-based implementation that offers the possibility of exploiting our method using CPU or GPU without the need to download, install and configure the required software and hardware.

3 The proposed solution

In this section we present our solution within three parts : the sequential approach, the GPU embedded solution, the cloud solution.

3.1 Sequential solution

Before presenting the cloud-based implementation, we start by describing the main steps of our left ventricle detection and tracking method. Our method disposes of two main steps (Fig. 1).

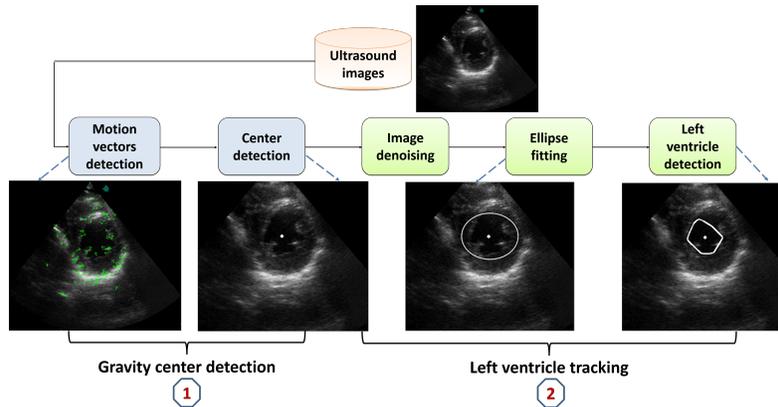


Fig. 1. Main steps of our method of left ventricle detection and tracking

1. **Gravity Center Detection:** this step consists of detecting the gravity center of the left ventricle, which allows to recognize and track the shape of left ventricle within the next step. This step is performed with three sub-steps: pre-processing, motion vectors detection and center detection as shown in Fig. 1.
2. **Left Ventricle Tracking:** the second step consists of exploiting the detected gravity center in order to detect and track the shape of left ventricle in real time. This step is composed of two sub-steps: approximation of left ventricle contour and its detection. These steps are detailed in our previous publication [10].

3.2 GPU-embedded solution

Despite the high accuracy of the proposed method, the computation time is so significant, which makes the approach not adapted for real time processing of echocardiography videos. The high increase of calculation time is due to three main reasons:

- Many image processing algorithms applied on each echocardiographic video frame
- The intensive computation of image pre-processing and left ventricle tracking algorithms
- The high definition of ultrasound images that requires more computation.

To overcome this constraint, we developed a cloud-based solution that allows to exploit GPUs without the need to download, install and configure the related hardware and software. On the one hand, we ported and parallelized the high intensive steps (pre-processing, motion vectors and gravity center detection) of our method on GPU using the API CUDA, while the less intensive steps (contour approximation and left ventricle detection) remain implemented on CPU

[10]. For more details about the GPU implementations, we refer readers to our previous implementations [11] [12] [13] [14].

We adapted our implementation of exploiting mobile graphic cards "NVIDIA TEGRA TK1" in parallel. The exploitation allows to reduce to the energy consumption by maintaining the benefit of exploiting the computing units of GPU in parallel [16]. The TEGRA TK1 GPU device allows to develop embedded applications that includes NVIDIAs highest performing and all the modern graphics and compute APIs (OpenGL, DirectX, etc.). The TEGRA TK1 GPU provides 4 CPU cores (64 bits), 192 CUDA cores and a memory of 2 GO.

As result of this exploitation, we obtained accurate left ventricle tracking in real time and with a low energy consumption [17] as shown in Table 1. Notice that the obtained performance within the embedded GPU are lower (24.1 fps) that those obtained within Physical GPU (28 fps) [10] since the latter disposes of more CUDA cores and memory size. Despite this, the TEGRA TK1 GPU offered a real time processing for the left ventricle detection and tracking method. However, the embedded GPU solution offers an energy efficient solution since it consumes 10 times less that the normal GPU solution [10]. The use of GPUs within cloud platforms allows to get fast results without the need to configure and install software and hardware, which represents a complicated task for users or even developers. Experimentations are conducted using a set of 11 normal and 17 disease hearts echographic video sequences. The sequences are provided from the department of cardiology at the Hospital of Tlemcen in collaboration with Dr. Abbou ⁴. The ultrasound videos present different resolutions:

- Low resolution videos: 320×240
- Medium resolution videos: 640×480
- High resolution videos: 1058×794

	<i>Masuda et al.</i>	<i>Takahashi et al.</i>	Our method
Success rate	85 %	88 %	93 %
Automatic	Semi-automatic	Semi-automatic	Fully-automatic
Computing unit	CPU	CPU	GPU
Frame rate	4 fps	4 fps	24.1 fps

Table 1. Our method vs literature algorithms using a normal case video (1058 × 794)

⁴ Department of cardiology, Tlemcen University Hospital, Tlemcen, Algeria

3.3 Cloud solution

In our cloud solution, we have used a virtual machine (VM) that allows the access to our application of left ventricle detection and tracking. The related web address is: <https://toolbox.media-process.com/>. Users can test and try the application within this address. This responsive application is developed using the PHP and Bootstrap that offered a multi-platform website running even on mobile devices (smartphone, tablet, etc.). The access to our cloud-based application is secured within HTTPS protocol. Otherwise, we used the Docker framework in order to provide a multi-user exploitation, different users can run the same application simultaneously (Fig. 2). Notice that Docker container is an open source software platform that can be used to manage the process of development. Its main benefit is to package applications in containers, allowing them to be portable among any system running the Linux operating system (OS). With Docker [15], we generated and configured an image called nvidia-docker-image including the operating system (Ubuntu and the required libraries : OpenCV and CUDA that allows to exploit NVIDIA GPUs.

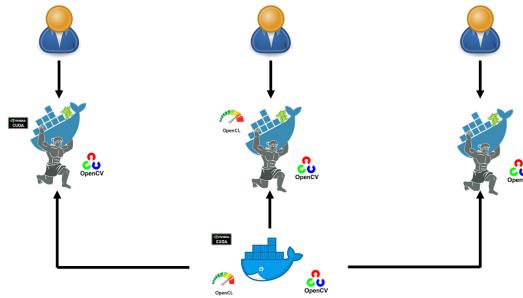


Fig. 2. Multi-user execution

The process of exploiting the cloud application is summarized within four steps (Fig. 3):

- Web selection: users can select the application of left ventricle detection and tracking within the platform.
- Input parameters uploading: users can provide the input parameters (the ultrasound video images) that will be sent to the web server.
- Cloud-based execution: the cloud platform generates the related nvidia-docker images with all the parameters in order to execute the application. In case we have many users, the platform creates a container for each user.
- Results presentation: at the end of the process of left ventricle and tracking, all the containers will be removed by the cloud platform and show the results to the user. Fig. 3 illustrates the general architecture of our cloud platform.

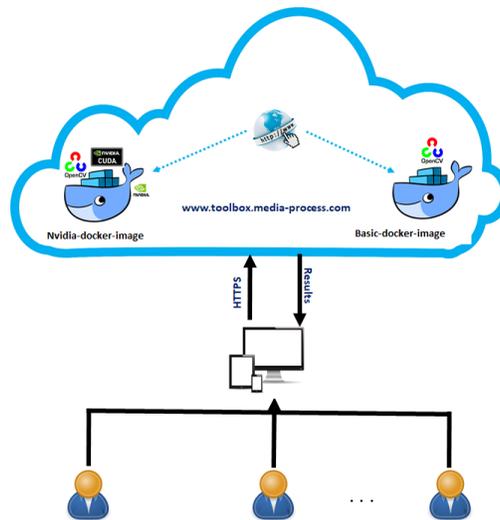


Fig. 3. The cloud platform architecture

4 Conclusion

In this paper, we proposed an accurate, automatic, GPU-embedded and cloud method for left ventricle segmentation and tracking in real time. The method is based on optical flow estimation for detecting the left ventricle center and convex hull for its tracking. Experimentations showed promising results (93 % of accuracy) thanks to the parallel exploitation of GPUs computing units that offered a real time processing even when treating high definition echographic videos. The accelerated GPU treatment (24 fps), which was 6 times faster than the CPU version (4 fps). The use of embedded GPUs (Tegra TK1) allowed to reduce the energy consumption, while the use of cloud platforms allowed to facilitate the exploitation of our method since users do not need to download, install or configure the related software and hardware. As future work, we plan to validate our method by the use larger data-set of ultrasound images, which allows to get apply a training of data for better accuracy. We plan also to improve the performance of results visualization in the cloud using a double buffering system. We plan also to apply our method for detecting and tracking left ventricle with larger data sets of 3D ultrasound videos.

References

1. X. Papademetris, A. J. Sinusas, D. P. Dione, and J. S. Duncan, Estimation of 3d left ventricular deformation from echocardiography. *Medical Image Analysis*, vol. 5, no. 1, pp. 17 – 28, 2001.

2. K. Masuda, R. Takahashi, T. Yoshinaga, and S. Uchibori, *Elucidation of intersection distribution in motion vectors from successive echocardiograms and its application for heart disease recognition*, World Congress on Medical Physics and Biomedical Engineering, Munich, Germany, pp. 572–574, 2009.
3. M. Suhling, M. Arigovindan, C. Jansen, P. Hunziker, and M. Unser, “Myocardial motion analysis from b-mode echocardiograms,” *IEEE Transactions on Image Processing*, vol. 14, no. 4, pp. 525–536, April 2005.
4. X. Papademetris, A. J. Sinusas, D. P. Dione, and J. S. Duncan, “Estimation of 3d left ventricular deformation from echocardiography,” *Medical Image Analysis*, vol. 5, no. 1, pp. 17 – 28, 2001. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1361841500000220>
5. T. Dietenbeck, M. Alessandrini, D. Barbosa, J. Dhooge, D. Friboulet, and O. Bernard, “Detection of the whole myocardium in 2d-echocardiography for multiple orientations using a geometrically constrained level-set,” *Medical Image Analysis*, vol. 16, no. 2, pp. 386 – 401, 2012.
6. Sidi Ahmed Mahmoudi, Michal Kierzynka, Pierre Manneback, and krzysztof Kurowski. Real-time motion tracking using optical flow on multiple gpus. *Bulletin of the Polish Academy of Sciences, Technical Sciences*, 62(1):139–150, 2014.
7. Sidi Ahmed Mahmoudi, Erenca Ozkan, Pierre Manneback and Suleyman Tosun Taking Advantage of Heterogeneous Platforms in Image and Video Processing. In *High-Performance Computing on Complex Environments*, pp 429-449 , 2016.
8. R. B.-I. Calasanz, “Towards the scientific cloud workflow architecture,” *5th International Workshop on ADVANCEs in ICT Infrastructures and Services (ADVANCE’2017)*, 2017.
9. H. M. Agrawal. Cloudcv: Large-scale distributed computer vision as a cloud service. *Mobile cloud visual media computing*, pp. 265–290, 2015.
10. S. A. Mahmoudi, M. Ammar, G. L. Joris, and A. Abbou, “Real time gpu-based segmentation and tracking of the left ventricle on 2d echocardiography,” in *Bioinformatics and Biomedical Engineering*. Cham: Springer International Publishing, 2016, pp. 602–614.
11. Sidi Ahmed Mahmoudi and Pierre Manneback. Multi-gpu based event detection and localization using high definition videos. In *International Conference on Multimedia Computing and Systems (ICMCS)*, pages 81–86, 2014.
12. Sidi Ahmed Mahmoudi and Pierre Manneback. Multi-cpu/multi-gpu based framework for multimedia processing. In *Computer Science and Its Applications, vol. 456, pp 54–65*, 2015.
13. Paulo R. Possa et al. A new self-adapting architecture for feature detection. In *22nd International Conference on Field Programmable Logic and Applications (FPL)*, pages 643–646, 2012.
14. Sidi Ahmed Mahmoudi and Pierre Manneback. Efficient exploitation of heterogeneous platforms for images features extraction. *3rd International Conference on Image Processing Theory, Tools and Applications*, pages 91–96, 2012.
15. Merkel D. Docker: lightweight linux containers for consistent development and deployment. In *Linux Journal*, 239, 2014.
16. Sidi Ahmed Mahmoudi et al.. Real-Time GPU-Based Motion Detection and Tracking Using Full HD Videos. In *International Conference on Intelligent Technologies for Interactive Entertainment*, pp 12-21, Belgium , 2013.
17. Sidi Ahmed Mahmoudi et al. Towards a smart selection of resources in the cloud for low-energy multimedia processing. In *Concurrency and computation practice and experience*, 2017.

**Web-based application for accurately classifying cancer type from
microarray gene expression data using a support vector machine
(SVM) learning algorithm**

Shrikant Pawar^{1,2*}

¹Department of Computer Science, Georgia State University, 34 Peachtree Street, 30303, Atlanta, GA, USA. ²Department of Biology, Georgia State University, 34 Peachtree Street, 30303, Atlanta, GA, USA.

Correspondence Email: *spawar2@gsu.edu

Abstract:

Intelligent optimization algorithms have been widely used to deal complex nonlinear problems. In this paper, we have developed an online tool for accurate cancer classification using a SVM (Support Vector Machine) algorithm, which can accurately predict a lung cancer type with an accuracy of approximately 95 percent. Based on the user specifications, we chose to write this suite in Python, HTML and based on a MySQL relational database. A Linux server supporting CGI interface hosts the application and database. The hardware requirements of suite on the server side are moderate. Bounds and ranges have also been considered and needs to be used according to the user instructions. The developed web application is easy to use, the data can be quickly entered and retrieved. It can be easily accessible through any web browser connected to the firewall-protected network. We have provided adequate server and database security measures to the system, with no client software need and supporting major operating systems (Windows, Linux and OSX) with a dedicated industry standard server. The developed application will help

researchers to utilize machine learning tools for classifying cancer and its related genes.

Availability: The application is hosted on a personal Linux server and can be accessed at:

<http://131.96.32.330/login-system/index.php>

Keywords:

Cancer; Microarray; Support Vector Machine (SVM).

Introduction:

Many machine learning algorithms such as random forest, k-nearest neighbor, neural network, and SVM (Support Vector Machine) have been applied in pattern classification. SVM has distinctive advantages in handling data with high dimensionality and a small sample size. Expression levels of genes can be analyzed using microarray experiments, in which the RNA isolated from different tissues is labeled and hybridized to the arrays [1]. The expression levels of treatment sample can be compared to control sample to understand the differences in gene expression levels amongst two treatments. Microarrays can be of two type: single channel where only one dye (red or green) is used, and the two channel where two dyes are used, one for control (CY5) and other for treatment (CY3). Different array spots give different fluorescence intensity values for each gene, which can be analyzed with different bioconductor package in R to perform normalization and statistical analysis.

Normalization is essential prior to making comparisons amongst biological replicates. It is required as the RNA used for hybridization can be of different quantities or there can different labeling conditions for different probes or the scanning of expression levels may be biased. Normalization essentially adjusts the expression levels of probes to reference probes [1]. Figure no 2 is the one channel microarray lung cancer data after normalization. A significant change in expression levels with correction of standard deviation and variance within and between samples

is observed. We read four lung cancer samples in R and measured standard deviation and variance on raw data to compare it with an RMA normalized data, a significant variance stabilization was obtained in RMA normalized data.

A SVM is a machine learning technique which categorizes new samples based on supervised learning of training data. A hyperplane is defined which forms a minimum distance in training sample which is used for classification [2]. SVM makes classifications based on non-probabilistic binary linear classifier [3]. It has been shown that SVM's can significantly increase accuracy compared to traditional query refinement schemes [3]. Although, several groups have shown importance of machine learning tools in classifying big data, there still remains a constant need to develop a user friendly tool for implementing this algorithm on big data. We have developed this functional tool, which can input a one channel microarray data and accurately predict the cancer type using machine learning technique. Although, we only have incorporated one algorithm in this article, developments are in progress for utilizing several others (linear regression, logistic regression, naïve bayes, random forest, K-means, K-nearest neighbor).

Materials and Methods:

a. Data Collection:

A set of 103 one channel microarray lung cancer samples (GPL590) were downloaded from Gene Expression Omnibus (GEO) database. Same number of normal lung tissue microarray data accompanied this cancer data. The work flow for this analysis is depicted in Figure 1. The downloaded raw data is then directed towards normalization for variance stabilization.

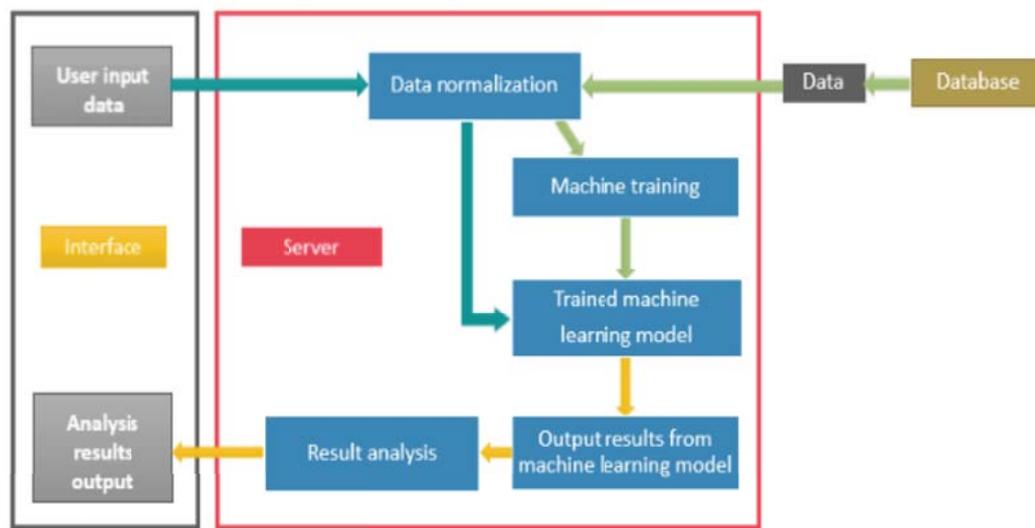


Figure no 1: depicts the work flow for this project. User inputs a microarray data, it is downloaded in database, normalized, analyzed by SVM, and the results of which displayed on an HTML page.

b. Normalization:

Normalization was performed using two techniques mas5.0 and RMA normalizations. We settled on RMA normalization because of its faster speed and improved stabilization quality [4]. It initially reads the affymetrix data using AFFY package in R language to perform RMA normalization with subsequent extraction of expression levels using ‘exprs()’ function. Since it’s a non-logarithmic transformed data, a logarithm to base 2 transformations is performed for scaling results.

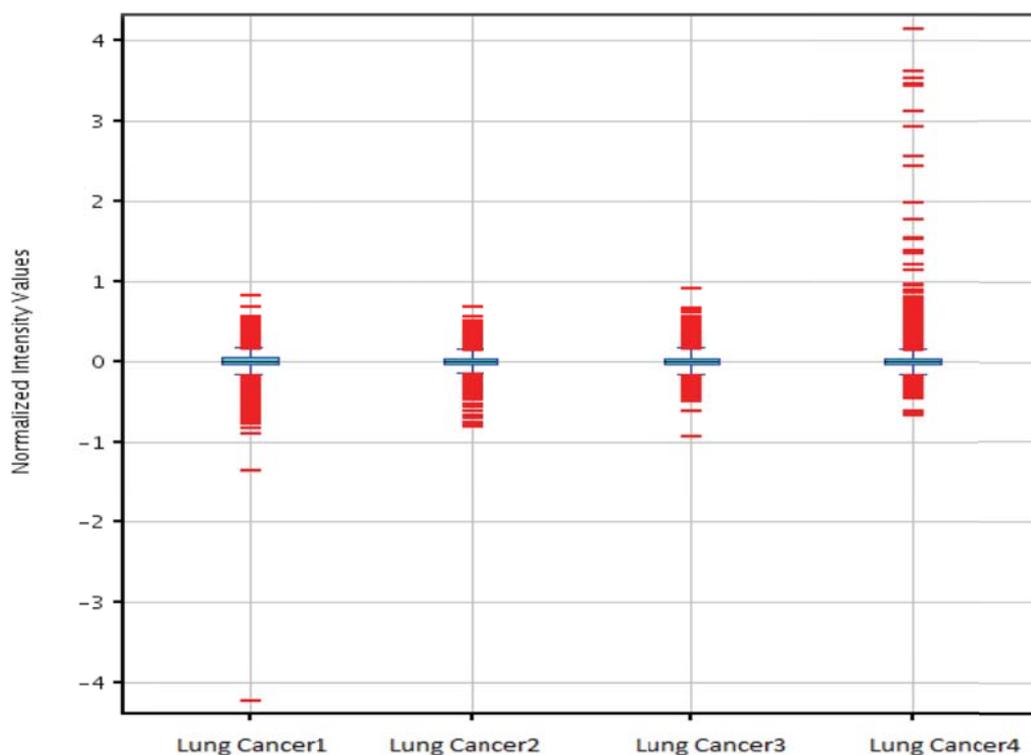


Figure no 2: Four lung cancer samples read in R, box plot depicts amount of standard deviation and variance stabilized by RMA normalization.

c. Machine Learning Analysis:

From 'sklearn.svm' python module penalty parameter C of the error term was set true ($C=1.0$), size of the kernel cache was 200 MB ($cache_size=200$), all classes are supposed to have weight one. The 'balanced' mode uses the values to automatically adjust weights inversely proportional to class frequencies in the input data ($class_weight=None$), $coef0$ is an independent term in kernel function and is only significant in 'poly' and 'sigmoid' ($coef0=0.0$), $decision_function_shape$ returns a one-vs-rest ('ovr') decision function of shape ($n_samples, n_classes$) as all other

classifiers, or the original one-vs-one ('ovo') decision function of 'libsvm' which has shape $(n_samples, n_classes * (n_classes - 1) / 2)$ (`decision_function_shape=None`), degree of the polynomial kernel function was set to 3 (`degree=3`), kernel coefficient for 'rbf', 'poly' and 'sigmoid'. The kernel type used is 'rbf' with maximum iterations of -1 (`max_iter=-1`), the probability estimates were enabled prior to calling fit (`probability=True`), `random_state` is the random number generator, in this case random number generator is the 'RandomState' instance used by 'np.random' (`random_state=None`), shrinking heuristic was set to true (`shrinking=True`), tolerance for stopping criterion was set to 0.001 (`tol=0.001`), verbose output was disabled (`verbose=False`), which takes advantage of a per-process runtime setting in 'libsvm' [5]. Classification is done using SVM. These are other set of supervised learning methods useful with high dimensional spaces where dimensions are greater than number of samples [6]. SVM is a memory efficient algorithm and uses subset of training points in the decision function [6]. Initially a training set of 103 cancer samples was trained to perform predictions on query data.

d. Setting up a Server Space and developing a Graphical User Interface (GUI):

GUI was developed by importing module 'Tkinter'. A 200x100 dimension frame was made for uploading files from the user. Server was developed with Python using 'BaseHTTPServer' module, other classes like 're', 'os', 'Classifier', 'Constants', 'normalize' were declared in main class 'MyHandler'. Three main methods are declared, `def 'do_GET(self)'` responding to a GET request, `def 'do_POST(self)'` calls normalization and classifier function and `def 'predict_data(self, test_data)'` responds to a GET request. `Def 'get_html_response_after_prediction(self, result, fileName)'` method writes the results on HTML page after predictions have been made. 'BaseHTTPServer' module creates and listens at the HTTP socket, 'Re' module accepts regular expression, 'Os' module allows to interface with operating system. Class 'constants' defines

HOST_NAME = '10.241.213.126', PORT_NUMBER = 9000 and paths for trained data. Based on the user specifications, we chose to write this suite in Python CGI and based on a MySQL database. It enables the set-up of single or a multi-users access controls. A Linux server hosts the application and database. It has been developed on a Mac OS X operating system using MySQL as the relational database management system and Python as the scripting language. The hardware requirements of suite on the server side are moderate. The server we utilized is GSU Orion with CentOS 6.7 64-bit, 6x IBM System x3850 x5, Intel Xeon Processor E7-4850, 4 CPUs (10 cores per CPU), 2.0 GHz processors, 512 GB RAM and 2 TB of scratch storage for jobs. The database has a column text-field table which is updated interactively with the user. Further, the uploading of information is standardized as certain parameters like date has to be inputted in specific format only which helps in retrieval process. Checking of data type (float, integer, text, Boolean) and dates is important before putting into database. Errors with dates or invalid parameters or wrong data type cause a halt of workflow. Bounds and ranges have also been considered and needs to be used according to the user instructions.

e. Security:

We have implemented strong security measures for authentication of SQL server. Kerberos protocol uses a number of encrypted messages to authenticate SQL server and the passwords are not passed across the network. Authentication is more reliable and managing it can be reduced by leveraging active directory groups for role-based access to SQL server. The sysadmin (sa) account is vulnerable when it exists unchanged so we have disabled the sa account on the SQL server instance. We chose to give options of complex passwords for sa and all other SQL-server-specific logins on SQL server and checked in the 'Enforce password expiration' and 'Enforce password policy' options for sa. We haven't allowed to explicit grant control server permission because

logins with this permission get full administrative privileges. For permissions to users, an built-in fixed server roles and database roles or creating own custom server roles and database roles can be achieved. Guest user exists in every user and system database, which is a potential security risk in a lock down environment because it allows database access to logins who don't have associated users in the database. We have restricted this access and also accesses to user and system stored procedures. Furthermore, we have used common specific TCP ports (excluding 1433 and 1434) instead of dynamic ports. SQL server browser service is only running on SQL servers, and secure SQL server error logs and registry keys using NTFS permissions are utilized as they can provide great deal of information about the SQL server instance and installation.

Results and Discussion:

An HTML page is created to provide options to the user for uploading data shown in Figure no 3. Once the user uploads a .CEL format file it calls functions in server, performs machine learning training and testing and uploads results on same HTML page as shown in Figure no 4. A sample lung cancer. CEL file (GSM159355) was uploaded by user as a query sample and the prediction analysis gave a prediction of 94.88 % accuracy of being a lung cancer type as shown in Figure no 4.

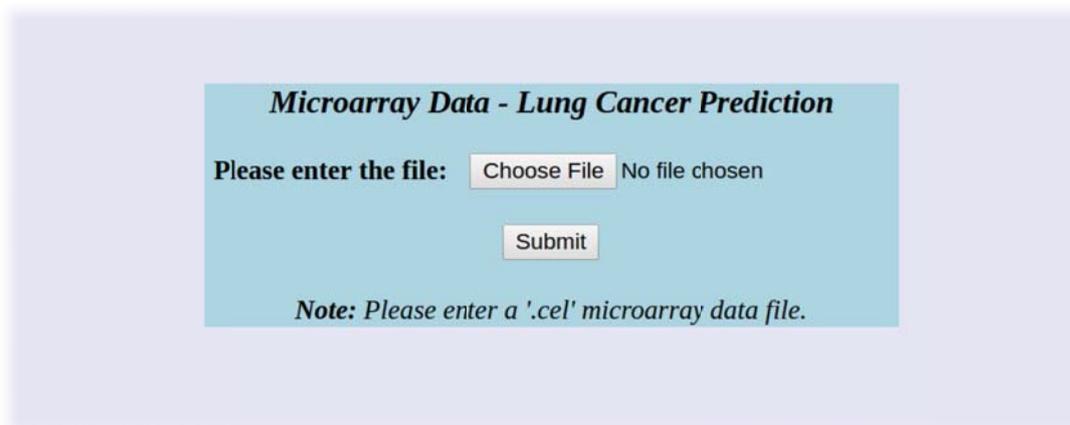


Figure no 3: HTML page provided for the user to upload the data.

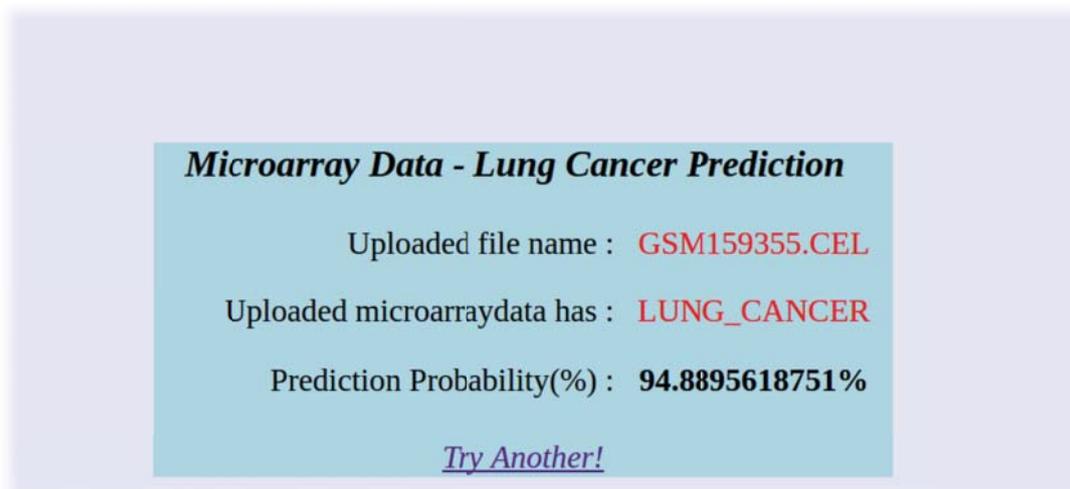


Figure no 4: Prediction analysis gave a prediction of 94.88 % to be a lung cancer type.

Conclusions:

Expansion of training set is needed to improve prediction accuracies. Also increasing cancer types (breast, brain, liver etc.) can make this program usable for different groups. Enhancement of the current pipeline is needed in terms of incorporation of other machine learning

algorithms. The current pipeline is a skeleton for predicting unknown genes involved in different cancer types using machine learning tools.

Funding:

No external funding for developing this suite has been utilized.

Competing interests:

The author declares no competing interests.

Author's contributions:

SP designed the study and critically revised the manuscript. SP developed, tested the software and also did the setup of SQL database and server space.

Acknowledgements:

Support from the Georgia State University Information Technology Department (GSU IT) for server space is gratefully acknowledged.

References:

1. Quackenbush J. (2002). Microarray data normalization and transformation. *Nature Genetics*, 496 - 501. doi:10.1038/ng1032.
2. OpenCV. (2014). Introduction to Support Vector Machine. Retrieved from http://docs.opencv.org/2.4/doc/tutorials/ml/introduction_to_svm/introduction_to_svm.html.
3. Ben-Hur, Asa, Horn, David, Siegelmann, Hava, and Vapnik, Vladimir (2001). "Support vector clustering". *Journal of Machine Learning Research*, 2: 125–137.

4. Laurent Gautier, Leslie Cope, Benjamin Milo Bolstad, and Rafael A. Irizarry (2003). Affy - an r package for the analysis of affymetrix genechip data at the probe level. *Bioinformatics*.
5. Alex J. Smola, Bernhard Schölkopf (2004). "A Tutorial on Support Vector Regression". *Statistics and Computing archive*. p. 199-222; 14(3).
6. Scikit-learn developers. (2014). Support Vector Machines. Retrieved from <http://scikit-learn.org/stable/modules/svm.html>.

Interaction between *Aedes aegypti* CPB1 and viral proteins

LÓPEZ- Edgar**†, BARRIENTOS- Carolina*

Laboratorio de Química Médica y Quimiogenómica, Universidad Veracruzana, Veracruz, México.*Abstract**

The chikungunya (CHIKV), zika (ZIKV) and mayaro (MAYV) viruses are etiological agents of emerging tropical diseases that represent a public health problem. Infection mechanisms occur through the *Aedes aegypti* vector where a series of viral proteins, such as Dengue virus E protein (DENV), interact. Thus, the identification of surface proteins similar to this, allows a better understanding of the mechanisms of internalization in their respective vector; which in the future will facilitate the generation of new therapeutic agents.

The objective of the present work is to analyze, by in silico methods, DNA proteins similar to the DENV E protein, where the organisms of interest are CHIKV (E1), ZIKV (E) and MAYV (P130) and to propose interaction sites between its capsid proteins and the CPB1 integral protein of the vector. Clustal Omega, I-TASSER, ClusPro online servers and the Protein Data Bank (PDB) database were used for this purpose. As well as the bioinformatic tools TMPRED and BLAST.

The binding sites, potential interacting amino acids, types of existing bindings and their type of environment were elucidated. This study allowed to determine an initial in silico antecedent, with which the same problem can be addressed, but with in vitro or in vivo studies.

Docking, *Aedes aegypti*, Dengue, Zika, Chikungunya, Mayaro, CPB1.

Introduction

Diseases that are transmitted by vectors, especially mosquitoes, are among the main causes of morbidity and mortality in humans. *Aedes aegypti* and *Aedes albopictus* are vectors that can infect human viruses, generating (to name a few diseases) dengue hemorrhagic fever (DHF) and dengue shock syndrome (DSS), emerging diseases worldwide transmitted by insects that threaten a third of the human population. During the bite of the female mosquito the person infected by the virus usually contracts it, becoming a carrier, where after several physiological events it is internalized in its midgut. Finally, the virus colonizes the salivary glands of the vector where it can now infect another person after feeding on their blood. (1-3)

That said, it was thought that the internalization method should be shared among different viruses, so the following research question was posed: is there structural convergence between viral proteins whose vector is *Aedes aegypti*?

Over the years, bioinformatics has generated great theoretical contributions that have subsequently been tested both *in vitro* and *in vivo*. That is why it is currently considered an important antecedent for any biological research.

Our work reflects the possibility of interaction of the selected proteins with the protein carboxypeptidase 1 (CPB1) in *Aedes aegypti*. CPB1 is located in the intestinal cells of the dipterid, which is experimentally defined as participating in the internalization of DENV, where the E protein of this pathogen is anchored to domain II of it. (1). This work postulates molecular interactions between proteins that are structurally similar to the E protein of DENV, so, we believe that its mechanism of internalization can be carried out by the same receptor protein in *Aedes*, we speak of CPB1.

Proteins of pathogens such as DENV and ZIKV (Flavivirus), CHIKV and MAYV (Togavirus) were evaluated. Where the inclusion criteria were: to be superficial, structurally conserved domains and sequential homology with the E protein of DENV.

Methodology

Homology of sequences by the algorithm of close neighbors.

Initially, the possibilities of analyzing and structurally comparing proteins from different viral organisms were considered, with the purpose of identifying if in addition to "homology" there was functional similarity. (4) However, the point of convergence selected for the

analysis of the mentioned viral proteins was the fact of owning the same vector (*Aedes aegypti*). It is known that these viruses use the aforementioned dipterous as a transport, for which reason they considered the possibility that they use similar internalization mechanisms.

Once the organisms of interest were defined, they were analyzed using the technique of "homology of sequences by algorithms of close neighbors". The product obtained was a phylogenetic tree. Which is composed of nodes and branches; These nodes can represent either an individual, a species, or a higher grouping and are therefore widely termed taxonomical units. In this case, the terminal nodes represent the species of analysis and are the operational taxonomic units (Otu). The ordering of the nodes determines the topology of the tree and describes how the lineages have diverged in the course of evolution. The branches of the tree represent the amount of evolutionary divergence between two nodes in the tree and can be based on different measurements.

A tree is completely specified by its topology and the set of all the lengths of the edges. (6)

Sequential analysis of the proteins of interest with Clustal Omega.

Once the proteins of interest for each of the organisms were identified, the similarity of their sequences was analyzed. This was done through the "Clustal Omega" server (<http://www.ebi.ac.uk/Tools/msa/clustalo/>).

When the homologies were obtained respectively for each protein, the next step was to individually analyze different characteristics such as the existence of transmembrane domains, using the TMPRED tool (http://embnet.vital-it.ch/software/TMPRED_form.html).

A BLAST of the E1 protein of DENV was made in the PDB and NCBI database obtaining no crystallizable version. Therefore, we proceeded to use bioinformatics tools such as the I-TASSER server (<http://zhanglab.ccmb.med.umich.edu/I-TASSER/>) which allowed us to make a structural prediction of the mentioned protein.

Molecular docking and search for interacting sites.

Once the generated model was obtained and selected, we performed a molecular Docking using the Clus-Pro server (<https://cluspro.bu.edu/home.php>). At the end of this process we observe and analyze the molecular couplings, selecting the models generated with structurally more stable scores (Balanced, Electrostatic, Hydrophobic and VdW + Elec).

Following the process, the software Swiss-PDBViewer v. 4.1.0 in each one of the proteins of interest were

introduced, in order to observe their supposed interacting sites.

Organism	Protein	ID / Location
DENV	E	None / I-TASSER
MAYV	P130	Q8QZ72 / UniProt
CHIKV	E1	3N42 / PDB
ZIKV	E	None / I-TASSER
Aedes a.	CPB1	None / I-TASSER

Table 1 Location of viral proteins. It shows the IDs in different databases respectively for each protein of interest. "None" refers that there is no crystallographic structure of the protein, for which models were generated by homology.

Results

Homology of sequences by the algorithm of close neighbors.

The diagram shows the relationship between the capsid protein sequences of some viruses, where the grouping of the proteins coming from different serotypes of DENV is clearly observed, immediately to this it bifurcates having on the one hand ZIKV with a value of 0.429 of its most immediate node and the CHIKV and MAYV proteins that are part of another branch with a distance of 0.064 and 0.495, respectively. Figure 1.

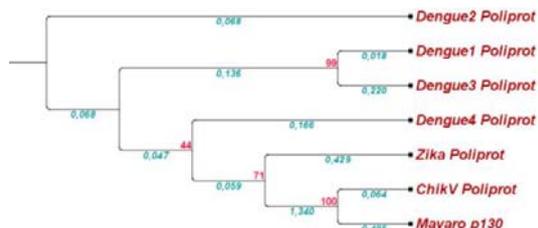


Figure 1 Phylogenetic tree of viral structural proteins. In blue refers phylogenetic distance between the sequences of the proteins to be studied and in red the percentage of times that the calculation generated this same results.

Sequential analysis of the proteins of interest with Clustal Omega.

The server was used introducing the sequence of each of the proteins of interest, being the E1 proteins of CHIKV, p130 of MAYV (Togaviridae family), E of DENV and ZIKV (Family Flavoviridae). With what was obtained the

following table shows the percentage of homology among all these. Figure. 2

	1	2	3	4	5	6	7
1.DENV1	100						
2.DENV2	72.04	100					
3.DENV3	79.17	66.67	100				
4.DENV4	69.06	69.33	65.48	100			
5.MAYV	17.15	17.52	17.95	18.61	100		
6.ZIKV	57.75	54.50	-nan	56.25	17.65	100	
7.CHIV	16.58	17.60	-nan	18.93	60.32	-nan	100

Figure 2 Homologous-sequential analysis of the proteins of interest. Alignments of biologically significant multiple sequences of amino acid sequence were produced. Evolutionary relationships are seen in percentage.

Molecular docking and search for interacting sites.

Complementing the results, we used the TMPRED server (http://www.ch.embnet.org/software/TMPRED_form.html), with the idea of splicing the possible transmembrane sites with the interacting assumptions obtained from the previous study. Figure 3

Obtaining significantly positive sites in the first 25 aa (being able to be a signal peptide), as well as approximate amino acids at 150, 225, 320 and 400. And sites statistically less significant at 150-200 and 350.

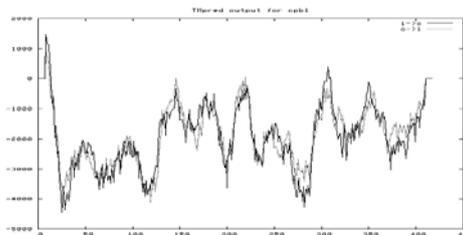


Figure 3. Possible transmembrane sites of CPB1. The algorithm is based on statistical base analysis, a database of natural transmembrane proteins. The prediction is made using hydrophobicity matrices.

We used standard E protein in DENV, which Hong-Wai T. and Vinod R.M.T in 2014 reported their interacting amino acids with CPB1 protein in Aedes aegypti. We managed to reproduce these same results in our work, apart from the fact that in parallel we obtained new ones with the different proteins mentioned. The joining energies being the following: DENV (-203.3), MAYV (-157.3), ZIKV (-210.4) and CHIKV (-247.0).

The results of homology of sequences by means of the algorithm of close neighbors revealed the similarity between the analyzed proteins, existing between each one of the sequences "possible common ancestors", which

explains the possibility of having similar functions, such as the inter-nalization to the Aedes enterocyte in the same way, that is, through CPB1.

We were able to reproduce the previously published results on the interaction between the E protein of DENV and CPB1 in Aedes, which confirms the good manipulation of the bioinformatic data.

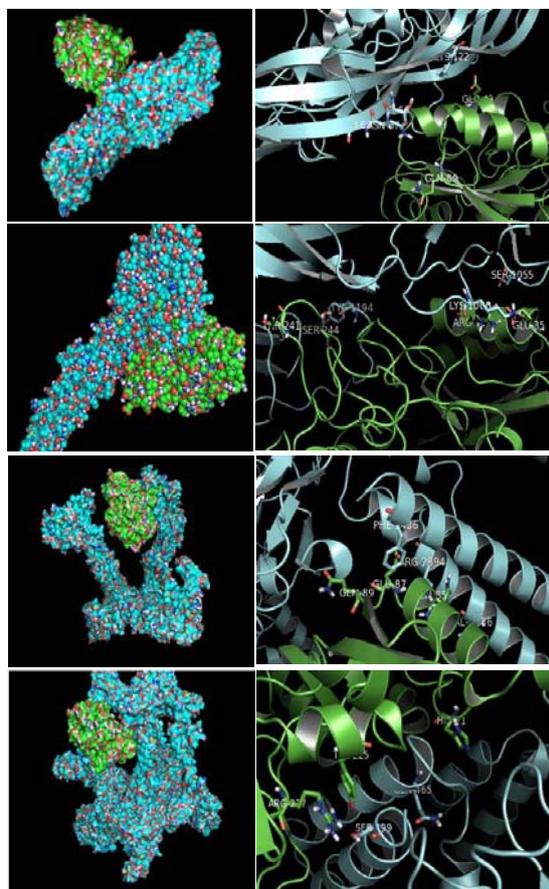


Figure 4 Molecular docking of viral proteins vs CPB1 of Aedes aegypti. In green CPB1 is shown and in blue the respective proteins of the different viruses.

Viral Protein	aa CPB1		aa Viral Protein	
	E (DENV)	ARG,	89,	THR,
PHE		90	ASN, THR	
GLU		43	LYS	122

D
E
N
V

M
A
Y
V

Z
I
K
V

C
H
I
K
V

	VAL, LEU	91, 65	VAL	251
P130 (MAYV)	GLU, ARG	35,42	LYS	1060
	ASN	39	SER	1055
	SER	244	TYR	1194
E1 (CHIKV)	GLN	89	ARG	2162
	GLU	87	PHE	2436
	GLN	85	ARG, VAL	2894, 2890
E (ZIKV)	HIS	21	---	---
	TYR	225	GLN	465

Table 2 Proposal of interacting amino acids. List of interacting aa in CPB1 and in the viral proteins evaluated.

Discussion.

We analyzed the interacting sites of CPB1 with the different proteins, obtaining that between the E protein of DENV and the E1 protein of CHIKV there are very similar couplings, as well as the p130 protein of MAYV and the E protein in ZIKV.

Therefore, the importance of the prediction of transmembrane helices of CPB1, which demonstrates the possibility of multiple binding sites is noteworthy. In addition, the existing interactions between the viral proteins and CPB1 conserve in all cases hydrophobic and polar characters, which are indispensable in the stability and viability of the proposed models.

Conclusions

The E protein in DENV is coupled to CPB1 in a similar way in terms of interaction type compared to the other protein models reported here. However, the binding sites were not homogenous, suggesting multiple binding sites in CPB1. The next step will be to evaluate the dimeric or trimeric conformations (depending on the case) of the different viral proteins, with the idea of simulating in a more realistic way the existing interactions in nature. The binding sites reported here may be used in the future for the development of CPB1 blockers, thus avoiding the excessive propagation of said viruses, and in turn improving public health strategies.

In silico methods provide several advantages for the resolution of biological problems, because through them it is possible to observe molecularly a macrophenomeno, in addition to reducing costs (with the use of free servers and academic softwares) and time in basic science.

of Sciences of the United States of America, 97(15), 8392–8396.

References

1. Hong-Wai Tham 1, V. R. (2014). CPB1 of *Aedes aegypti* Interacts with DENV2 E Protein and Regulates Intracellular Viral Accumulation and Release from Midgut Cell. *Viruses*. 6(12): 5028–5046.
2. Izabela A. Rodenhuis-Zybert, J. W. (2010). Dengue virus life cycle: viral and host factors modulating infectivity. *Cellular and Molecular Life Sciences*. 67(16):2773-86.
3. Kabra SK, J. Y. (1999). Dengue hemorrhagic fever: clinical manifestations and management. *Pediatr*, 66: 93-101.
4. Maher A. Nouredine, P. Y.-J. (2005). Genomic Convergence to Identify Candidate Genes for Parkinson Disease: SAGE Analysis of the Substantia Nigra. *Movement Disorders*. v.40(8).
5. Óscar Coltell, M. A. (2008). Integration of clinical and biological data in clinical practice using bioinformatics. *Rev Méd Chile*. 136(5):645-52.
6. Woese, C. R. (2000). Interpreting the universal phylogenetic tree. *Proceedings of the National Academy*

Aging Effect on the Pelvis to Head Attenuation of Upper Body Rotation during Walking

Hyeong-min Jeon¹, Jae-hoon Heo¹, Eui-bum Choi¹ and Gwang-Moon Eom¹

¹ Konkuk University, Chungju, Chungbuk 274-78, Korea
gmeom@daum.net

Abstract. The abstract should summarize the contents of the paper in short terms, i.e. 150-250 words.

Keywords: Upper Body, Rotation, Attenuation, Walking, Aging.

1 Introduction

Balance control is one of the main tasks of human walking, which is a complex coordinated movement of whole body [1]. Sensory functions related to balance control includes somatosensory (proprioception), vestibular, and visual system [2]. The head, containing two of these (vestibular and visual system), needs to be stabilized in space to provide a steady internal reference [3,4]. Vestibular system plays an important role in balance control by providing rotational speed and linear acceleration to the central nervous system [2,5]. In normal walking, the head is stabilized in translation and rotation to provide a stable reference for vision [4-1]. Vestibulo-ocular reflex enhances visual performance by providing stable focus on the retina, and rotational speed is the main information for the reflex [5].

Stabilization of the head was shown to be provided by attenuation (pelvis to head) of acceleration in the upper body, transmitted from the lower limbs [3,6,7]. Though, the stabilization of head includes both the rotational speed and linear acceleration, the attenuation has not investigated for the rotational speed. Age-related decline in balance control may be related to the head stabilization [4]. Therefore, this study aims to investigate the pelvis to head attenuation of upper body rotation in young and elderly subjects, so as to elucidate the aging effects.

2 Method

Fifteen young and fifteen elderly men participated in this study. Inclusion criteria were no occurrence of brain disease or orthopedic injury in recent one year.

Three-axis wireless gyroscopes (Triago, Delsys, Natick, MA) were attached on four back bony surfaces of body: on the head (occipital bone), the shoulder (C7), the thorax (T10) and the pelvis (sacrum). Subjects were requested to walk on a level surface

(Distance: 8m / width: 1.5m) at their own comfortable walking speeds. Two trials of walking were measured and analyzed.

3 Result

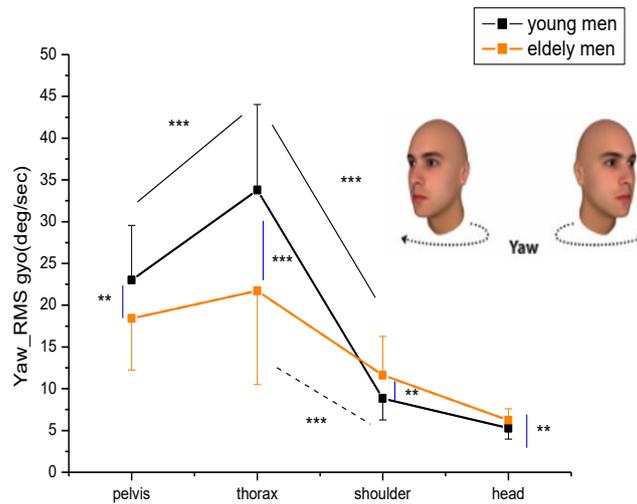


Fig. 1. Age-group comparisons of rotation speed concurrent with post-hoc pairwise comparisons of age-position interactions. (**: $p < 0.01$, ***: $p < 0.001$)

Two-way ANOVA showed interaction of two factors (age and sensor location) on the dependent variable (RMS angular velocity) in yaw direction ($p < 0.01$). Post-hoc pairwise comparison results are shown in Fig. 1. In the yaw direction, the pelvis and thorax angular velocity was greater in the young ($p < 0.01$). Thorax to shoulder attenuation of angular velocity was significant in the young but not in the elderly. As a result, the shoulder and head angular velocity was greater in the old ($p < 0.01$).

4 Discussion

The smaller rotation speed at the pelvis and thorax of elderly subjects (compared to the young; Fig. 1) may be related to reduced walking speed. The elderly showed reduced attenuation of thorax to shoulder attenuation, which resulted in greater rotation speed at the head (Fig. 1). The decline of upper body attenuation in the elderly may suggest altered coordination strategy in axial rotations [8, 9] and/or reduced flexibility of upper body.

Acknowledgement

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education and science. (2017R1A2B2010062, 2015M3A9D7067390).

References

1. Woollacott, MH., Pei-Fang, T.: Balance control during walking in the older adult: research and its implications. *Phys Ther* 77(6), 646 (1997).
2. Im, S.: Anatomy and Physiology of Balance. *Brain & NeuroRehabilitation* 6(2), 47-53 (2013).
3. Jeon, H., Kim, J., Kwon, Y et al.: Upper body accelerations during locomotion in different age groups and genders. *Journal of Mechanics in Medicine and Biology* 17(7), 1740026-37 (2017).
4. Paquette, C., Paquet, N., Fung, J.: Aging affects coordination of rapid head motions with trunk and pelvis movements during standing and walking. *Gait Posture* 24(1), 62-69 (2006).
- 4-1. Liu-Ambrose T, et. al. Resistance and agility training reduce fall risk in women aged 75–85 with low bone mass: a 6-month randomized, controlled trial. *J Am Geriatr Soc* 52, 657–65 (2004).
5. Park, H.: Vestibulo-Ocular Reflex for High Frequency and Acceleration Head Rotation. *Korean J Otolaryngol* 49, 356-65 (2006).
6. Mazzà, C., Iosa, M., Picerno, P et al.: Gender differences in the control of the upper body accelerations during level walking. *Gait Posture* 29(2), 300-303 (2009).
7. Kavanagh, JJ., Barrett, R., Morrison, S.: Upper body accelerations during walking in healthy young and elderly men. *Gait Posture* 20(3), 291-298 (2004).
8. Cavanaugh, JT., Shinberg, M., Ray, L et al. : Kinematic characterization of standing reach: comparison of younger vs. older subjects. *Clin Biomech* 14(4), 271-279 (1999).
9. Swinnen, E., Baeyens, J., Pintens, S et al.: Walking more slowly than with normal velocity: The influence on trunk and pelvis kinematics in young and older healthy persons. *Clin Biomech* 28(7), 800-806 (2013).

Potentiometric Screen-Printed Sensors for Wireless Monitoring of Wound Healing

Andrzej Peplowski¹✉, Daniel Janczak¹, Łukasz Górski², Michał Zbieć³, Dariusz Obrębski³, Małgorzata Jakubowska¹

¹ - Department of Microtechnology and Nanotechnology, Faculty of Mechatronics, Warsaw University of Technology, Warsaw, Poland

² - Department of Analytical Chemistry, Faculty of Chemistry, Warsaw University of Technology, Warsaw Poland

³ - ICs and Systems Design Department, Institute of Electron Technology, Warsaw, Poland

✉ - corresponding author, e-mail: a.peplowski@mchtr.pw.edu.pl

6th International Work-Conference on Bioinformatics and Biomedical Engineering, Granada, Spain, April 25-27 2018

Abstract: In this work we present a novel potentiometric cells printed directly on a commercial dressing patch. These cells can be used for measuring of the environment pH, an important indicator in the wound healing process. Fabrication of sensing and reference electrodes employing printing pastes is described and wireless reading of the sensor is presented that does not require energy supply beyond read-out smartphone device. Presented sensor exhibits linear response in pH range 5.5-7.9 with potential slope $-16.7 \text{ mV}\cdot\text{pH}^{-1}$.

Keywords: *printed electronics · potentiometric sensors · textronic sensors · wound healing · dressing · smart bandage*

1. Introduction

Wound treatment is a part of routine processes in modern healthcare, concerning both posttraumatic damage and surgical tissue incisions. Although majority of these injuries heal properly, incidental transition to chronic, non-healing wound due to infections and other factors causes increased patient mortality and also a drastic rise of treatment costs [1]. In the United States, chronic wound treatment costs 25 billion dollars [2] while in Europe 25-40% of hospital beds are being occupied by patients with wounds [1]. Thus, continuous monitoring of the wound condition is required to minimize patients suffering and lower the socioeconomic costs. Ideally, the wound status should be assessed without removing of the dressing – to prevent contamination and in real time – to start appropriate treatment, when infection symptoms arise.

One of the well documented indicators of the infection is pH of the wound [3], which may vary from 5.5 to 8.5 [4]. In this work potentiometric measurement was chosen since previously achieved results with printed pH sensors suggested that this approach may be more advantageous in terms of fabrication simplicity [5,6], measurements range [7] and flexibility.

2. Experimental

2.1. Fabrication of sensors

Using commercially available adhesive dressing patch was employed as a substrate for screen printing of the potentiometric sensors. Pastes used for printing were described previously by the authors [8]. Briefly, 8 wt% solution of poly(methyl methacrylate) (PMMA) in diethylene glycol butyl ether acetate was used as pastes' carrier. As a functional phase, 5 wt% of graphene nanoplatelets (GNP) (thickness 5nm, diameter 25 μ m) and 60 wt% of ruthenium (IV) dioxide submicron powder was used for working electrodes, while 85 wt% silver microflakes (diameter 2-4 μ m) were used as a filler for electrical connections and reference electrodes. UV-curable 5018G paste purchased from DuPont (USA) was used for insulating layers.

First, insulating layer was screen printed to prevent short circuit of electrodes and cured for 10 min. under UV lamp. Next, silver contacts for electrodes were printed and cured in 120°C for 30 min. On the contacts, RuO₂/GNP paste was deposited and cured in 120° for 30 min. to yield sensing electrodes. On this structure, another insulating layer was printed as previously, leaving only working and reference electrodes uninsulated. As the last step, 0,25M FeCl₃, 0,2M HCl solution was drop-casted on the reference electrode area to form Ag/AgCl layer. Schematic structure and photograph of the fabricated sensor are shown in Fig. 1.

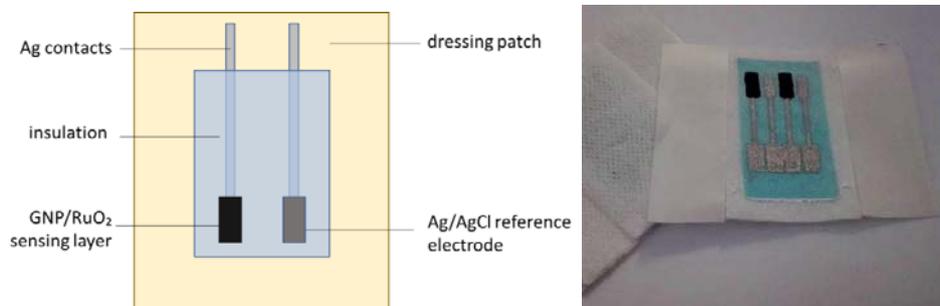


Fig. 1. Left: schematic of the pH sensor; right: example of the fabricated sensor on a dressing patch.



Fig. 2. NFC read-out device and sensor connected to the smartphone application for pH measurement.

2.2. Wireless connection device

The described readout system for pH printed electrodes was designed in a form of a small PCB containing complete electronic circuit of the NFC tag. The high miniaturization level was possible due to the utilization of the NT3H2211 (NTAG) integrated circuit from NXP. It combines the full functionality of the NFC passive tag, memory accessible via I²C interface and requires the minimum number of external passive components for proper operation. The printed board contains the ATmega88V 8-bit microcontroller from the AVR family (Microchip, formerly Atmel), performing the measurements cycles using its built-in 10-bit ADC. The microcontroller and remaining components of the system are supplied from the energy harvesting circuitry of the NTAG, so the whole sensor set-up does not require any internal energy source.

3. Results

On the fabricated sensor were drop-casted 1 M phosphate buffer solution (PSB) with pH ranging from 5.7 to 7.9. Potentials were registered using smartphone application 5 min. after casting of the solution. Obtained results are shown in Fig. 3. Slope calculated from linear regression was $-16.7 \text{ mV}\cdot\text{pH}^{-1}$, which is significantly below Nernstian response. It may be ascribed to partial swelling of the dressing material despite the insulation. Thus, previously swelled analyte solutions might affect sensor indications for subsequent samples. This problem, however, should not play major role in wound monitoring, since wound pH changes in longer time periods.

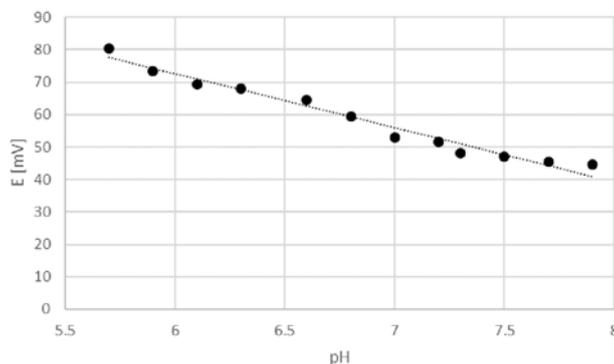


Fig. 3. Calibration curve of the pH sensor printed on a wound dressing patch, measurements carried out in PBS solution using NFC read-out application.

4. Conclusions

Flexible potentiometric pH sensor printed on a regular wound dressing patch was calibrated using wireless communication device and a smartphone application. Range of pH measured with described setup corresponds to the pH of properly as well as infected wounds. Thus, onset for development of feasible, fully functional tool for monitoring of wound healing process is confirmed, which can contribute to more effective and cheaper treatment of injured patients.

References

1. Posnett, J., Gottrup, F., Lundgren, H., & Saal, G. (2009). The resource impact of wounds on health-care providers in Europe. *Journal of wound care*, 18(4), 154-154.
2. Services UDoHaH. (2004) Guidance to surveyors for long term care facilities. Guidance to Surveyors for Long Term Care Facilities on World Wide Web URL: <http://www.cms.hhs.gov//>
3. Schneider, L. A., Korber, A., Grabbe, S., & Dissemond, J. (2007). Influence of pH on wound-healing: a new perspective for wound-therapy?. *Archives of dermatological research*, 298(9), 413-420.
4. Dargaville, T. R., Farrugia, B. L., Broadbent, J. A., Pace, S., Upton, Z., & Voelcker, N. H. (2013). Sensors and imaging for wound healing: a review. *Biosensors and Bioelectronics*, 41, 30-42.
5. Guinovart, T., Valdés-Ramírez, G., Windmiller, J. R., Andrade, F. J., & Wang, J. (2014). Bandage-Based Wearable Potentiometric Sensor for Monitoring Wound pH. *Electroanalysis*, 26(6), 1345-1353.
6. Phair, J., Newton, L., McCormac, C., Cardosi, M. F., Leslie, R., & Davis, J. (2011). A disposable sensor for point of care wound pH monitoring. *Analyst*, 136(22), 4692-4695.
7. Schaude, C., Fröhlich, E., Meindl, C., Attard, J., Binder, B., & Mohr, G. J. (2017). The Development of Indicator Cotton Swabs for the Detection of pH in Wounds. *Sensors*, 17(6), 1365.
8. Janczak, D., Peplowski, A., Wroblewski, G., Gorski, L., Zwierkowska, E., & Jakubowska, M. (2017). Investigations of Printed Flexible pH Sensing Materials Based on Graphene Platelets and Submicron RuO₂ Powders. *Journal of Sensors*, 2017.

Graphene layers and coatings in biomedical applications

Lucja Dybowska-Sarapuk^{1✉}, Andrzej Kotela², Malgorzata Jakubowska¹

1 - Warsaw University of Technology, Faculty of Mechatronics, Department of Microtechnology and Nanotechnology, Andrzeja Boboli 8, 02-525 Warsaw, Poland

2 - Department of Orthopaedics and Traumatology, 1st Faculty of Medicine, Medical University of Warsaw, Lindleya 4, 02-005 Warsaw, Poland

✉ - corresponding author, e-mail: l.dybowska@mchtr.pw.edu.pl

6th International Work-Conference on Bioinformatics and Biomedical Engineering, Granada, Spain, April 25-27 2018

Abstract.

Graphene nanoplatelets are the carbonaceous material with unique properties, very often used in electronics, but recently more often used in medical applications. Graphene, as a variety of carbon, potentially harmless to humans, can be used in tissue and cellular engineering, in drug delivery or as an anti-bacterial material.

On the other hand, the ideal layer - coverage for bioapplications such as implantology should include the ability to block bacterial biofilm and ability to simulate tissue integration, good mechanical strength, biocompatibility and antibacterial properties [1]. Gallo et al. claim that there are no such layers and still is a real need to find material which although in part fulfills such requirements. In our work, we took the challenge and decided to prove that our graphene layers could be the perfect material for biological applications.

1 CYTocompatible MATERIALS – INKS

The first task was to formulate cytompatible graphene inks for printed electronics techniques, with functional phase in a form of graphene nanoplatelets, which are presented in Figure 1.

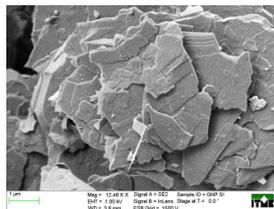


Fig. 1. SEM image of graphene nanoplatelets

It was very difficult because such bioinks must meet requirements from two completely different science fields: both engineering and biological. The ink cannot be toxic to the cells, the required condition is the biocompatibility. However, it is equally important that the ink has to be easily printable. Therefore, it must meet the rheological requirements such as suitable viscosity and surface tension [2].

We succeeded in producing inks that not only were cytocompatible and printable but also supported stem cell proliferation.. [3] The viability of stem cells is shown in Figure 2.

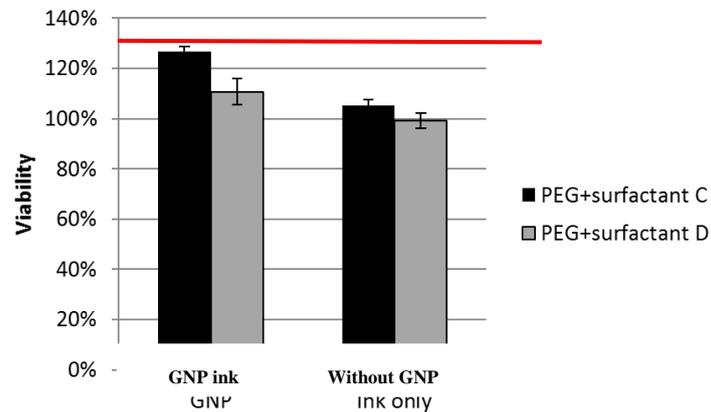


Fig. 2. Cytotoxicity of GNP coatings [3].

2 PRINTING METHODS OF LAYERS

Printed electronics techniques such as ink-jet printing and spray coating, shown schematically in the Figure 3 and Figure 4 are economically efficient way of producing homogenous layers and coatings with a good adhesion and different thicknesses.

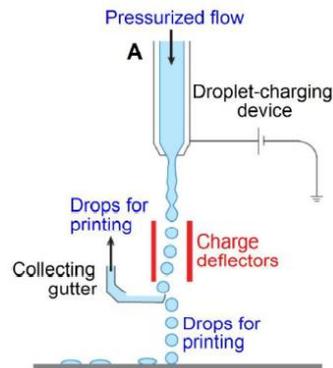


Fig. 3. Scheme of ink-jet printing technique [4].

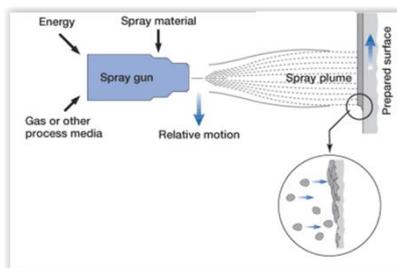


Fig. 4. Scheme of spray coating technique [www.kondex.com]

The advantages of these techniques include: cleanliness and simplicity of the process, high resolution and possibility of printing on different types of materials, including flexible substrates with various shapes (non planar). Our coatings were printed on various substrates including silicon substrates or Foley Catheters and characterized by very good mechanical properties [5], [6].

3 PREVENTING BIOFILM FORMATION

Biofilms play a huge role in a wide variety of microbial infections and therefore remain a significant source of morbidity and mortality in the world's population. Nowadays It is a challenge in founding a new, effective way of preventing biofilm formation. Therefore we have tested our graphene layers also for ability of effective blocking the bacterial biofilm [6]. SEM images of GNP layers coatings after incubation with *S. epidermidis* are shown on Figure 5.

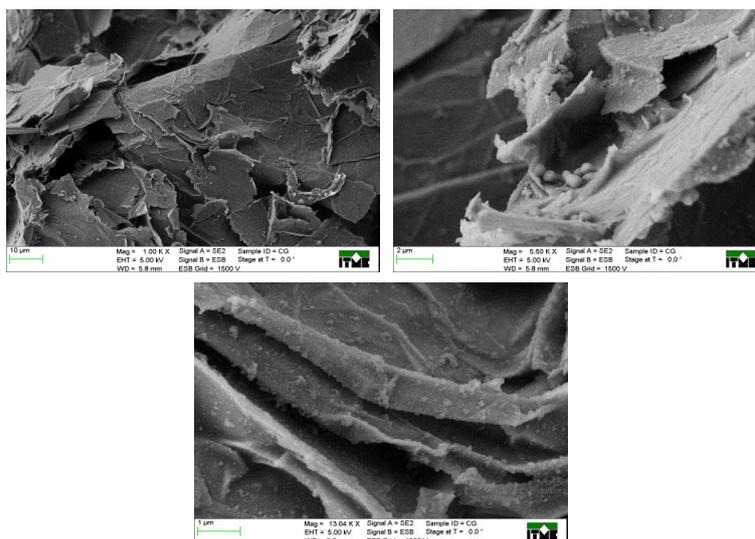


Fig. 5. SEM images of GNP coatings after cell culture [6]

The findings indicate the ability of graphene layers to prevent the formation of biofilm [5], [6].

Concluded, we are able to produce graphene layers that simultaneously meet requirements related to: cytocompatibility, biodegradability, the ability to stimulate tissue integration, the ability to block bacterial biofilm. Moreover these layers characterized by good conductivity, excellent mechanical strength and adhesion.. In the future, they may find a number of applications in medicine in the manufacture of medical tools such catheters or in coatings on bone or vascular implants.

4 References

- [1] J. Gallo, M. Holinka, and C. S. Moucha, Antibacterial surface treatment for orthopaedic implants, *Int. J. Mol. Sci.*, vol. 15, no. 8. 2014.
- [2] K. Woo, D. Jang, Y. Kim, and J. Moon, Relationship between printability and rheological behavior of ink-jet conductive inks, *Ceram. Int.*, vol. 39, no. 6, pp. 7015–7021, 2013.
- [3] L. Dybowska-Sarapuk, S. Rumiński, G. Wróblewski, M. Słoma, A. Młozniak, I. Kalaszczyńska, M. Lewandowska-Szumieł, and M. Jakubowska, “Aqueous biological graphene based formulations for ink-jet printing,” *Polish J. Chem. Technol.*, vol. 18, no. 2, pp. 46–52, 2016.

- [4] B. Derby, "Inkjet Printing of Functional and Structural Materials: Fluid Property Requirements, Feature Stability, and Resolution," *Annu. Rev. Mater. Res.*, vol. 40, no. 1, pp. 395–414, 2010.
- [5] Ł. Dybowska-Sarapuk, A. Kotela, J. Krzemiński, D. Janczak, M. Wróblewska, H. Marchel, P. Łęgorz, and M. Jakubowska, "Antibacterial activity of graphene layers," vol. 10031, p. 100310G, 2016.
- [6] Ł. Dybowska-Sarapuk, A. Kotela, J. Krzemiński, M. Wróblewska, H. Marchel, M. Romaniec, P. Łęgosz, and M. Jakubowska, "Graphene nanolayers as a new method for bacterial biofilm prevention: Preliminary results," *J. AOAC Int.*, vol. 100, no. 4, pp. 900–904, 2017.

Computational Modelling of the Possible Cancer Synergic Treatment Combining Oncolytic Treatment and Induced Immune Cytotoxicity

López-Valverde, José-Antonio

Seville University. Seville, Spain. joslopval14@alum.us.es

Abstract. Treating cancer with oncolytic viruses is one of the most remarkable emerging therapies against cancer, mostly because the improvement that it would suppose to the quality of patient's life. However, there is some controversial data about the interaction between this therapy and the immune system. In the following paper, a new insight about this interaction is proposed, using it as a mechanism of eliminating cancer cells. A cytotoxic T-CD8⁺ immune response will be used for it, using this immune reaction against cells that are infected by the viruses, producing a synergy in the treatment with the oncolytic activity of viruses. Computational modelling approaches are developed for testing this hypothesis and establishing the bases of this new field for improving the quality of oncolytic therapies, and cancer treatment in general.

Keywords: Cancer, oncolytic, virus, cytotoxic, immunotherapy, computational modelling.

1 Introduction

Developing new therapies and treatments for cancer is one of the biggest research areas at this moment [1]. This is due to the high necessity of new treatment implementation and the use of less aggressive therapies than chemotherapy, which along radiotherapy and surgery when they are possible, constitutes the first choice treatment in the majority of cancer cases [2]. More concretely, there are two areas with the biggest development: targeted therapy, based on the molecular singularities the malignant cells that are not presented on healthy cells, and immunotherapy, based on immune system resources (or resources that interact, modulate or mimic it) used to attack to the cancerous process [2].

In targeted therapy, one possibility oncolytic therapy, based on the use of genetically modified viruses that infects cancerous cells without infecting normal cells. They also take advantage of common cancerous cells mutations for replicating themselves, thus, the absence of this mutations in healthy cells prevents viruses to replicate in them. Therefore, oncolytic therapy is based on tumor cells suppression due to the cytopathic effect of the viruses, while normal cells are barely affected [3].

Nevertheless, there are some problematic aspects with the implementation of the therapy. One of the main controversial points is the virus-immune system interaction, because if it neutralizes the viruses, it would suppose the end of the treatment. However, not all the results of this interaction are negative. It has been demonstrated that immune system activation by the viruses can lead to a general activation state of the immune system, catalyzing the anti-tumor immune response [4, 5].

Also, the interaction with the viruses follow complex non-linear dynamics, related with other complex dynamics, so that's why there are several works on computational modelling of oncolytic therapy. Nevertheless, in the majority of models, the interaction with the immune system is not described [3].

In this paper, we aim to give a positive value to the virus-immune system interaction. More concretely, it's modelled what would happens if viruses with certain characteristics that cause the immune system to attack the infected cells (the cancerous ones) are used, thus causing a decrease of the infected population by the cytopathic effect of the oncolytic virus, and by the induced cytotoxicity when the immune system kills infected cells in attempting to eliminate the virus.

For this interactions to occur, it's necessary to have some conditions. For example, the virus must be CD8⁺ immunogenic predominantly, being this one the cytotoxic immune response. Also, the virus must be able to proliferate with the immune system, because if the virus gets neutralized in an early phase, timing could not be enough for the therapy to be effective.

In conclusion, the strategy is innovative: a synergy between classic oncolytic effects and induced cytotoxic response effects is proposed. *Ergo* the viruses are used like a bait for the lymphocytes to execute the tumor cells, and the *per se* anti-tumor immunity would co-exist.

2 Materials and procedures

The first step was defining the dynamics that were going to be studied, and the factors that influence in them. 5 dynamics will be considered: healthy cells (x), cancerous cells (y), infected cells (z), immune system cells (w) and viruses (v). This dynamics must follow mathematical behavior according to existing validated models.

Starting with the healthy cells (x), the dynamics will tend to equilibrium x_0 that represents the total amount of cells in a healthy adult individual [6]. This population decrease by the low affinity infection that suffers from the viruses, going each infected cell to the infected cells dynamics. This behavior is given by equation 1 with the following parameters (table 1):

$$\frac{dx}{dt} = r \cdot (x_0 - x) - \beta \cdot x \cdot v \quad (1)$$

Table 1. Parameters used on equation 1.

Parameter	Description	Value
x_0	Healthy cells initial population	$3.72 \cdot 10^{13}$ (cells)
r	Growth rate, healthy cells	0.1
β	Infectivity rate, healthy cells	10^{-18}
v	Viral population	Dynamic (viruses)

For the growth rate determination, the data provided in bibliography was used [7]. This data states that 50-70 billion cells are replaced daily, in a dynamic equilibrium state, but there should be taken into account that out of that equilibrium, cells reproduction is bigger.

The last member of the equation makes reference to the cells that get infected. A low infectivity rate is expected, given that this kind of cells hardly ever get infected by the viruses, due to the safety mechanisms explained before.

The cancerous cells population (y) increase by cell division, and decrease by infection and apoptosis. The behavior of this population is given by equation 2 and its parameters (table 2).

$$\frac{dy}{dt} = r' \cdot y \cdot \left(1 - \frac{y}{k}\right) - \alpha \cdot y - \lambda \cdot y \cdot v \quad (2)$$

Table 2. Parameters used on equation 2.

Parameter	Description	Value
y_0	Cancerous cells initial population	10^8 (cells)
r'	Growth rate, cancerous cells	0.5
k	Tumor carrying capacity	10^{11} (cells)
α	Apoptosis rate	10^{-2}
λ	Infectivity rate, cancerous cells	$1.8 \cdot 10^{-13}$

y_0 is the number of cells in a recently diagnosed tumor with 1 gram mass [8]. The growth is logistic with the growth rate r' , following previous models [5]. k represents a 1 kg tumor mass, quantity that it's not generally exceeded because of the patient dies, or the cancer suffers metastasis as a possible consequence of that carrying capacity [9].

Furthermore, there are two decrease factors, α with a value according to previous models [5] and λ , analogous to β but applied to cancerous cells. The cells that suffer infection are preferably cancerous cells, so that, λ is much bigger than β . Each healthy cell infected correspond with the infection of 100 000 cancerous cells.

On the other hand, there is the infected cells population (z), that increases by the infections, while decreases by the cytopathic effect of the virus, and by the cytotoxic

effect of the immune system. According to the previous considerations, equation 3 should be considered, with parameters on table 3:

$$\frac{dz}{dt} = \beta \cdot x \cdot v + \lambda \cdot y \cdot v - \varphi \cdot z - \mu \cdot w \cdot z \tag{3}$$

Table 3. Parameters used on equation 3.

Parameter	Description	Value
z_0	Infected cells initial population	0 (cells)
φ	Viral cytopathic effect rate	0.04
μ	Immune cytotoxicity rate	0.1
w	Immune system population	Dynamic (cells)

The cytopathic effect depends on the viral vector that it's used. In this case, a moderate cytopathic virus is considered, according to previous models [3, 5].

Other dynamic to model is the immune system cells (w). A system of equations, depending on the state of activation/inactivation of the immune system. The amount of viruses in the organism determines the use of one or the other equation. With the immune system not being active, the population tends to be in an equilibrium state. If it's active, it will suffer a clonal expansion (exponential growth) that is being countered in lesser way by apoptosis and cellular stress. There are equations 4 and 5, with parameters defined on table 4.

$$\begin{cases} \frac{dw}{dt} = \delta \cdot (w_0 - w) & v < \gamma \\ \frac{dw}{dt} = r'' \cdot w - o \cdot w - \frac{\sigma \cdot z}{w} & v > \gamma \end{cases} \tag{4, 5}$$

Table 4. Parameters used on equations 4 and 5.

Parameter	Description	Value
γ	Activating threshold	10^8 (viruses)
w_0	Immune system cells initial population	10 (células SI)
δ	Growth rate, inactive immune system	0.15
r''	Growth rate, active immune system	0.4
o	Apoptosis rate	10^{-4}
σ	Cellular stress rate	0.05

The threshold is established taking into account the viral weights [10, 11] and the quantity of viruses in a vaccine [12, 13]. In a vaccine there are about $2 \cdot 10^9$ viruses, quantity in excess for triggering the activation of immune system. That's why γ has to be smaller. w_0 is small because of the high specificity of the T-cell receptors, that are unique, although there are some lymphocytes that can respond to the same antigen with different affinity.

Furthermore, there's an apoptosis rate, that is similar to the one from normal population [7], and there's a cellular stress rate, meaning that the lymphocytes that suffer stress are no longer effective against their targets. Stress is inversely proportional to the quantity of effective lymphocytes, but directly proportional to the amount of infected cells: the more infected cells per lymphocyte, the more stress.

The last dynamic is the one corresponding to viruses, which population increase by self-replication them and decrease by the immune system action. Equation 6 states this behavior, with parameters on table 5.

$$\frac{dv}{dt} = v \cdot r''' \cdot \left(1 - \frac{v}{k'}\right) - \omega \cdot w \cdot v \quad (6)$$

Table 5. Parameters used on equation 6.

Parameter	Description	Value
v_0	Viral initial population	10^{11} (viruses)
r'''	Growth rate, viral population	1
k'	Viral population carrying capacity	10^{13} (viruses)
ω	Viruses elimination rate	$5 \cdot 10^{-4}$

The viral initial population is high because we want a generalized infection, that will follow a logistic dynamic with parameters according to bibliography [10-13]. The viral population decrease by the elimination of viruses by the immune system.

With the dynamics established, the model has been implemented using modeling and simulating software MATLAB, using the Euler simplified method for solving the differential equations.

3 Results and discussion

3.1 Model validation

For determining whether the model is valid or not, simulation approaches are made, and the results are compared with other models on the bibliography. Due to the lack of experimental data about this hypothesis, validation is only made by comparison with other models that were already experimentally validated, and by evaluating the physiological compatibility of the results obtained.

First of all, further understanding of the treatment of time measures is needed. It's expressed as arbitrary units because there are parameters that have not been experimentally determined yet, and they can change depending on the viral vector used, because of the interpersonal variability, etc. Changing this parameters proportionally would lead to having the same response to the treatment but in different amounts of time. Thereby, for not introducing subjective approaches about how much time would pass in the treatment response, arbitrary units are used, thus not decreasing the external validity of the modelling.

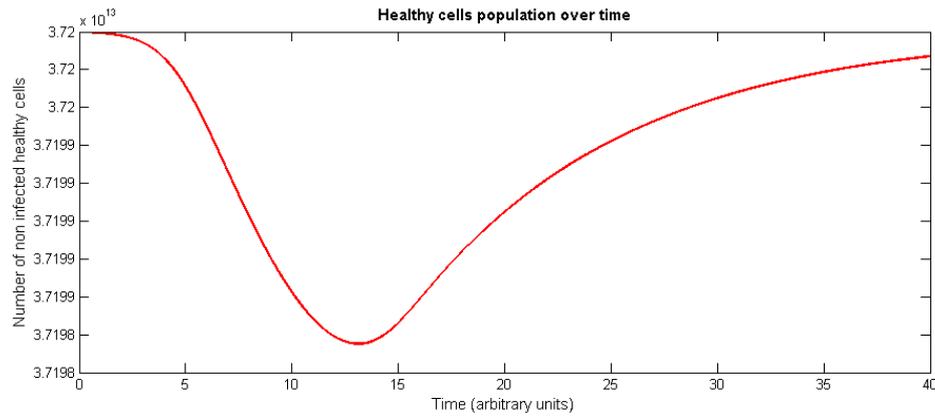


Fig. 1. Healthy cells population over time (40 arbitrary units).

Healthy cells population (Fig. 1) only suffers a slight decrease by the residual effect of infection by viral units. This effect is compensated by the return to equilibrium state after the therapy ends. Even so, more specific viral particles could be used characterizing the antigens by RNAseq or targeting multiple antigens (double targeting/multiple targeting) [2].

More attention should be given to the affected healthy cells, because if they all belong to the same set of cells, the damage to that cellular pool could be greater (although the number of cells affected is so low), constituting a significant damage.

The behavior shown in this dynamic match with the current physiological knowledge, and with already validated models [3, 5, 14].

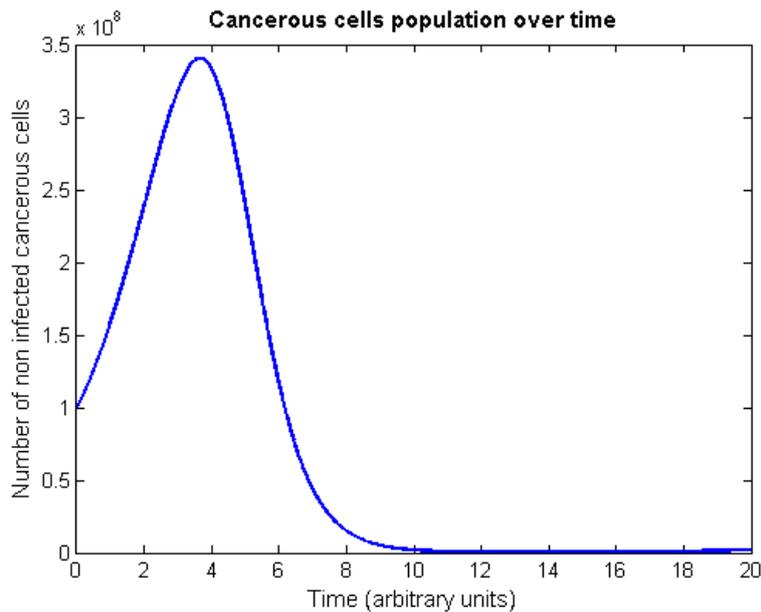


Fig. 2. Cancerous cells population over time (20 arbitrary units)

Cancerous cells population (Fig. 2) keeps growing in the presence of low amount of viruses (from time 0 to time 4 approximately). When the amount of viruses is bigger (time > 4), the cancerous cells population decrease.

However, this model supposes that the target is present in all the malignant cells, which is not always correct due to the tumor heterogeneity and the mutation capacity. Highly conserved structures should be chosen as targets. In addition, three-dimensional approaches are not considered into this model.

Anyway, the tumor dynamics fits into the expected responses, and fits with previously experimentally validated models [3, 5, 15]

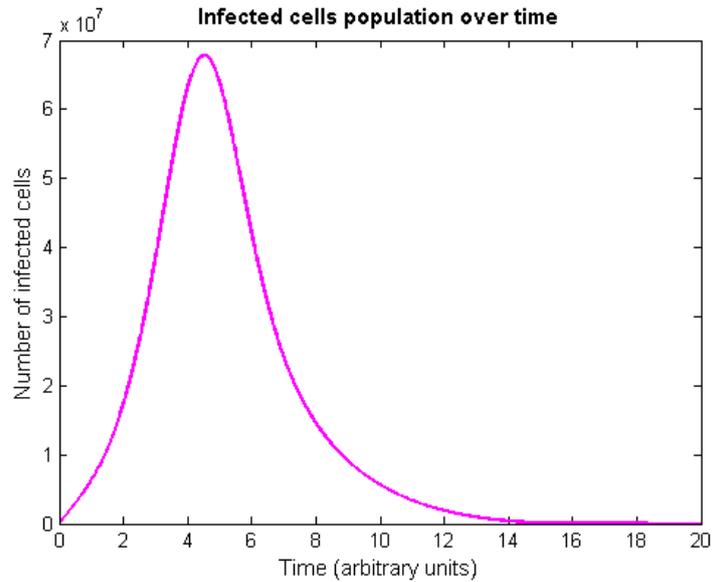


Fig. 3. Infected cells population over time (20 arbitrary units)

Infected cells population dynamics (Fig. 3) follow a bell-shaped distribution: it starts from 0, increase to a maximum while there is more infection than cellular de cease, and returns to 0 again when there is more cellular de cease than infection. This distribution is consequent with the rest of dynamics and with what would be expected in experimental validation.

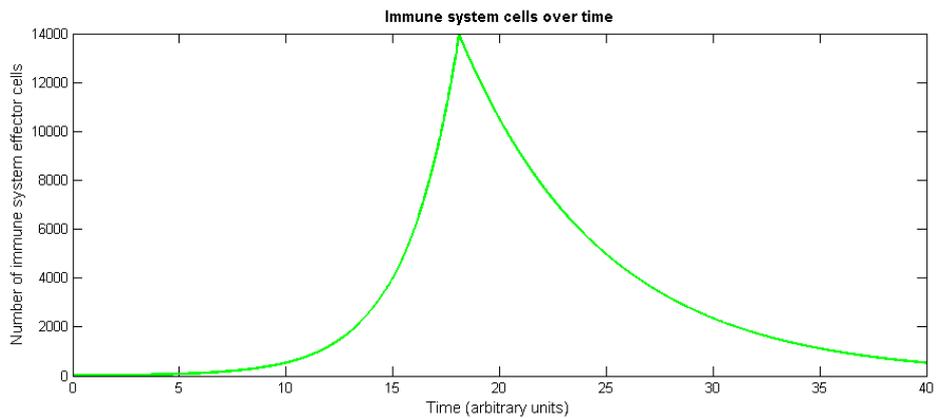


Fig. 4. Immune system effector cells over time (40 arbitrary units)

When talking about T-CD8⁺ effector cells (Fig. 4), an increase in their amount can be observed from the first moment ($t = 0$). This is because γ threshold is exceeded, causing the activation of clonal multiplication routes in certain immune system cells.

When the amount of viral particles is reduced, the population tends to an equilibrium state by immunomodulation. Although immune system memory is not included in the modelling, it's not important because re-inoculating the same viral particles is not pretended.

A future remark for this model could be the integration of the delay on the shut-down of proliferation after the number of viruses is lower than γ [16]. Also, probabilistic factors could be incorporated in parameters like σ , with a more accurate behavior as in reality: the more infected cells killed, the more probability of developing cellular stress.

The distribution is consequent with previously validated models [3, 5, 14].

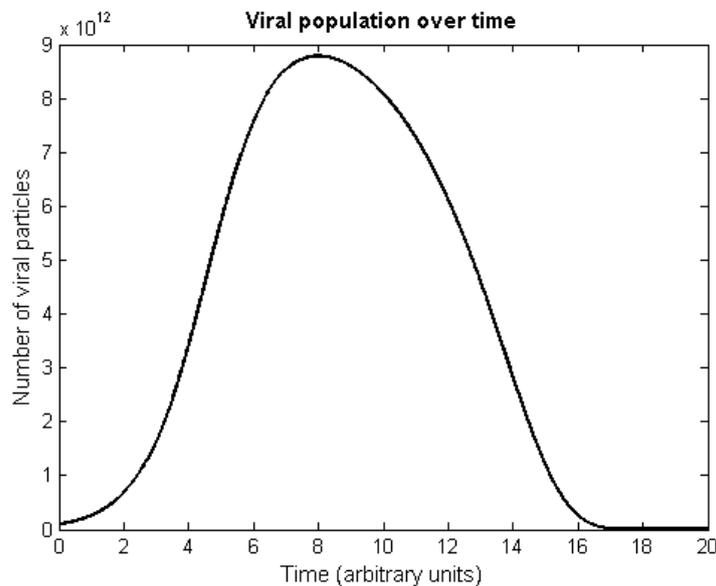


Fig. 5. Viral population over time (20 arbitrary units)

The viral population (Fig. 5) also follows a bell-shaped distribution, increasing when the viral replication is higher than the elimination by the immune system and vice versa.

A continuous growth in the population is modelled, which it does not have to be accurate with reality, because some viruses need to lyse the cells for being able to infect other cells, thus not contributing to the infective population. Figure 5 represents the concept of oncolytic therapy: a transient viremia with therapeutic purpose [17, 18].

3.2 Sensitivity analysis

A sensitivity analysis is performed, and it leads to the demonstration that the parameters in this model can be classified into two groups: sensitive parameters that affect the prognosis, and non-sensitive parameters, or parameters that despite they are sensitive, they are not affecting prognosis, only affecting to isolated dynamics. Examples of parameters from the first group are λ and ω , while γ or φ are representative parameters of the second group. In the figures 6 and 7, we shown examples.

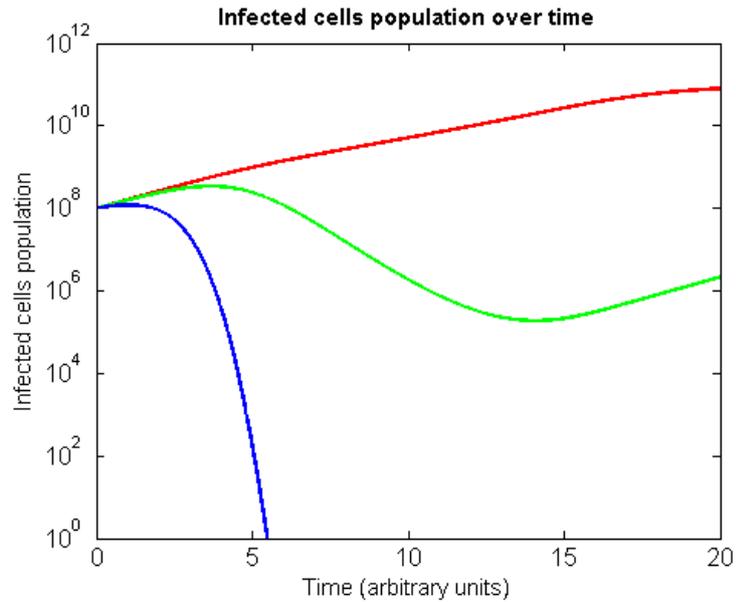


Fig. 6. Sensitivity of λ , a sensitive parameter that modify prognosis. Red, $\lambda=1.8 \cdot 10^{-14}$; Green, $\lambda=1.8 \cdot 10^{-13}$; Blue, $\lambda=1.8 \cdot 10^{-12}$.

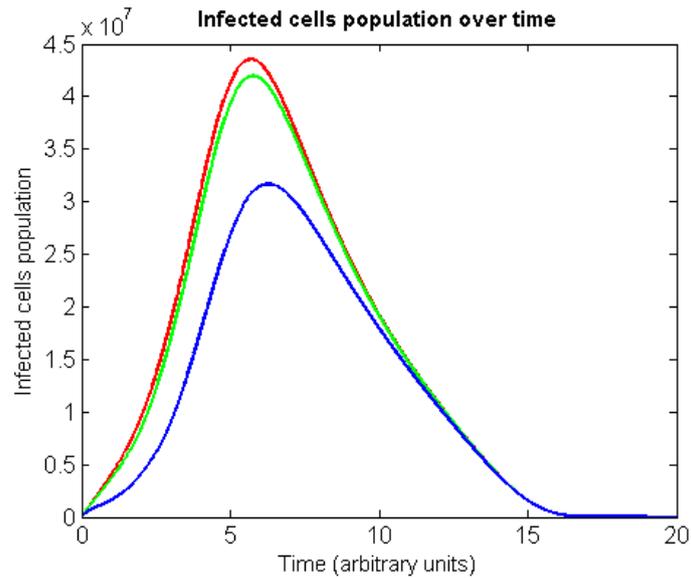


Fig. 7. Sensitivity of ϕ , a sensitive parameter that don't modify prognosis. Red, $\phi=0.04$; Green, $\phi=0.4$; Blue, $\phi=4$.

ϕ don't modify the prognosis at all because it's a relevant parameter to know how infected cells die, but it does not modify the amount of cancerous cells, only affect on how fast the cells die.

4 Conclusions

The results deliver support to the initial hypothesis, and reveal that the use of oncolytic therapy for inducing immunological cellular cytotoxic response is a new possible field to explore.

In short, the model is a good approach to the use of immune cytotoxic response by T-CD8⁺ lymphocytes as an instrument to improve the treatment, and the results define the basic guidelines to follow on this research area. The behavior of the computational model fits with other models results that have been experimentally validated.

Parameters determining is an essential future step to establish the order of the temporal magnitude, and the final characteristics in clinical practice. Also, there are some parameters that can be modified in experimental procedures for getting the desired prognosis. This is the versatility what constitutes a solid argument to continue investigating this treatment as possible future therapy of cancer.

Acknowledgements

I would like to thanks to all the professors on the “Biomedical Engineering” subject in the Basic and Experimental BSc in the University of Seville. Without the training they have given to me, this work could not be possible. Also, I would like to thank to my family and partners for making me feel encouraged about this project.

References

1. Quora, <https://goo.gl/3KEHVy>
2. National Cancer Institute (NIH), <https://goo.gl/EUgtXD>
3. Woodarz, D.: Computational modeling approaches to the dynamics of oncolytic viruses. *Wiley Interdiscip Rev Syst Biol Med.* 8(3), 245-252 (2016)
4. Lichty, B., Breibach, C., Stojdl, D.B.J.: Going viral with cancer immunotherapy. *Nature Reviews Cancer.* 14(3), 559-567 (2014)
5. Woodarz, D.: Gene therapy for killing p53-negative cancer cells: use of replicating versus nonreplicating agents. *Human Gene Therapy.* 14. 153-159 (2003)
6. Bianconi, E., Piovesa, A., Facchin, F., Beraudi, R., Frabetti, F., Vitale, L., et al.: An estimation of the number of cells in the human body. *Annals of Human Biology.* 40(6), 463-471 (2013)
7. Karam, J.A.: Apoptosis in Carcinogenesis and Chemotherapy. Vol 1. Springer (2009)
8. Del Monte, U.: Does the cell number 10^9 still really fit one gram of tumor tissue? *Cell cycle.* 8(3), 505-506 (2009)
9. Deisboeck, T.S., Wang, Z.: Cancer Dissemination: A Consequence of limited Carrying Capacity? *Med Hypotheses.* 69(1), 193-177 (2007).
10. Gupta, A., Akin, D., Bashir, R.: Single virus particle mass detection using microresonators with nanoscale thickness. *Applied Physics Letters.* 84 (2004)
11. Peplow, M.: Tiny scales weight virus. *Nature physics portal research highlights* (2004)
12. The Journal of Immunology. Concentration and Purification of Influenza Virus for the Preparation of Vaccines. *Journal of Immunology.* 50(5) (1945)
13. Centers for Disease Control and Prevention (CDC), <https://goo.gl/a6BpJZ>
14. Arias, C.F., Herrero, M.A., Acosta, F.J.: Population mechanics: A mathematical framework to study T cell homeostasis. *Sci Rep.* 7(1), 9511 (2017)
15. Arias, C., Herrero, M., Cuesta, J., Acosta, F., Fenández-Arias, C.: The growth threshold conjecture: a theoretical framework for understanding T-Cell tolerance (Figure 1). *R Soc Open Sci.* 7(2) (2015)
16. López, A.G., Seoane, J.M., Sanjuán, M.A.F.: A Validated Mathematical Model of Tumor Growth Including Tumor-Host Interaction, Cell-Mediated Immune Response and Chemotherapy. *Bull Math Biol.* 76(11), 2884-906 (2014)
17. Nemunaitis, J.: Oncolytic viruses. *Investigational New Drugs.* 17(4), 375-86 (1999)
18. Fillat, C.: Controlling Adenoviral Replication to Induce Oncolytic Efficacy. *The Open Gene Therapy Journal.* 3, 15-23 (2010)

Adapting and step by step refinement of the Vegetative Vigour Terrestrial Plant Test for assessing ecotoxicity of aerosol samples

Katalin Hubai¹, Eszter Horváth¹, Bettina Eck-Varanka¹, Gábor Teke², Ádám Tóth¹, Nóra Kováts¹,

¹ Institute of Environmental Science, University of Pannonia, Veszprém, H-8200, Hungary

²Elgoscár-2000 Environmental Technology and Water Management Ltd., Balatonfűzfő, H-8184, Hungary

Keywords: ecotoxicity, atmospheric aerosol, vegetative vigour tests, higher plants

The ecological effect of airborne pollutants have been extensively studied using plants as bioindicator organisms. Wide range of endpoints have been used in bioindicator studies (reviewed by Grantz et al, 2003). However, in case of biomonitoring studies, neither exposure nor environmental concentration of pollutants can be known, therefore no exact cause-effect relationships can be established.

Causality can more precisely be established in ecotoxicological tests, however, relatively few studies are available where ecotoxicological effect of air pollution have been assessed under controlled laboratory conditions (e.g. Daresta et al., 2015). Our goal was to evaluate the applicability of the No. 227 OECD GUIDELINE FOR THE TESTING OF CHEMICALS: Terrestrial Plant Test: Vegetative Vigour Test, to adapt it for assessing the ecotoxicity of water soluble compounds of aerosol samples and to carry out necessary refinements. The test assesses the potential effects of potentially toxic samples on higher plants by spraying the test substance on the leaves and above-ground portions of plants. In the series of our studies, composite winter aerosol samples were used. PM10 aerosol samples were collected in Budapest (Hungary), at the main observatory of the Hungarian Meteorological Service between 25.12.2014 and 06.01.2015.

In our first study (Kováts et al., 2017), *Cucumis sativus* L. was used as test plant. Plants were grown in commercial soil, in green house following the environmental condition prescribed by the Guideline. The protocol recommends only one treatment, at the beginning of the exposure, in our test a second and third test group was set which received a repeated spraying. The following layout was followed: Test group 1 (TG1): Treatment: Day0; Test group 2 (TG2): Treatment: Day0, Day8; Test group 3 (TG3): Treatment: Day0, Day8, Day15. The test was terminated on Day22. Fresh weight, shoot length were measured and foliar necrosis were recorded.

No necrosis was experienced in the control and in TG1, the average damage value was 3.56 ± 0.429 in TG2 and 5.83 ± 0.804 in TG3, showing a slight concentration/dose-dependent relationship. Significant differences of the shoot length were found between the tested plant groups (Kruskal-Wallis ANOVA $H=10.97$, $df=3$; $P=0.01$). No significant differences were found regarding the dry weights of tested groups (Kruskal-Wallis ANOVA $H=4.31$; $df=3$, $P=0.23$). Based on the results, appearance of visual symptoms such as necrosis proved to be a more sensitive end-point than growth inhibition.

In our second study the sensitivity of additional end-points were investigated, such as chlorophyll a and b, carotenoids and peroxidase (POD) activity, in comparison to the original, perhaps more robust ones. The ecotoxicity of aqueous extract of urban PM10 aerosol samples was assessed using *Lycopersicon esculentum* Mill. plants. Layout was the same as in the previous study.

Different end-points showed different concentration-effect patterns. Considering growth inhibition, shoot length and biomass showed different sensitivity: no statistically significant inhibition in case of shoot length, while the extract elucidated significant biomass reduction in TG2 (TG1 and control ANOVA $F=2.733$, $p=0.08$; TG2 and control ANOVA $F=3.957$, $p=0.0348$, Tukey HSD: Control -TG2: $p=0.029$).). Considering photosynthetic pigments, the concentration-effect relationship observed was a typical 'stimulative effect at low concentration, inhibition at higher concentration' pattern. This is a typical stressor-response pattern occurring when the sample contains nutrients which might mask the toxic effect at low concentrations.

Peroxidase activity (POD) proved to be the most sensitive end-point. POD activity shows a similar concentration-effect relationship as biomass: while no statistically significant effect can be established for TG1, TG2 shows significant response (ANOVA: $F=14.27$, $p=0.005$, Tukey HSD: Control - TG2: $p=0.004$).

Analytical measurements revealed the presence of micronutrients (most possibly being responsible for the stimulative effect). Experienced ecotoxic effect might be related to the presence of polyaromatic hydrocarbons (total PAH content was $1.5 \mu\text{g/l}$) and toxic heavy metal such as Al ($145 \mu\text{g/l}$) or Pb ($31.9 \mu\text{g/l}$).

Based on these studies, we can conclude that (1) The Guideline has proven applicable for assessing ecotoxicity of aerosol samples or airborne contaminants; (2) in order to mimic real-world exposure, repeated treatments are necessary and (3) the original protocol needs to be modified, including more sensitive end-points such as stress enzyme activity. As peroxidase activity is a relatively cost-effective end-point, it should be definitely included in the future applications of the Guideline, especially when deleterious effects of air pollution on plants is to be assessed.

Acknowledgement

This work was supported by the BIONANO_GINOP-2.3.2-15-2016-00017 project. The authors also acknowledge the financial support of Széchenyi 2020 under the EFOP-3.6.1-16-2016-00015.

Abril GA, Wannaz ED, Pignata ML (2014) Source characterization and seasonal variations of atmospheric polycyclic aromatic hydrocarbons at an industrial and semi-urban area through a local-scale biomonitoring network using *T. capillaris*. *Microchem. J.* 116:77–86. doi: 10.1016/j.microc.2014.04.008

Alkio M, Tabuchi TM, Wang X, Colón-Carmona A (2005) Stress responses to polycyclic aromatic hydrocarbons in *Arabidopsis* include growth inhibition and hypersensitive response-like symptoms. *J. Exp. Bot.* 56 (421):2983–2994

Daresta, B.E., Italiano, F., de Gennaro, G., Trotta, M., Tutino, M., Veronico, P. (2015.) Atmospheric particulate matter (PM) effect on the growth of *Solanum lycopersicum* cv. Roma plants. *Chemosphere* 119, 37–42.

Grantz DA, Garner JHB, Johnson DW (2003) Ecological effects of particulate matter. *Environ Int* 29:213–239. doi:10.1016/S0160-4120(02)00181-2

Günthardt-Goerg MS, Vollenweider P (2007) Linking stress with macroscopic and microscopic leaf response in trees: New diagnostic perspectives. *Environ. Pollut.* 147:467–488. doi:10.1016/j.envpol.2006.08.033

Jochner S, Markevych I, Beck I, Traidl-Hoffmann C, Heinrich J, Menzel A (2015) The effects of short- and long-term air pollutants on plant phenology and leaf characteristics. *Environ. Pollut.* 206:382–389. doi: 10.1016/j.envpol.2015.07.040

Kováts N, Horváth E, Eck-Varanka B, Csajbók E, Hoffer A (2017). Adapting the Vegetative Vigour Terrestrial Plant Test for assessing ecotoxicity of aerosol samples. *Environ Sci Pollut Res* 24:15291–15298. doi: 10.1007/s11356-017-9103-5

Pignata ML, Gudiño GL, Wannaz ED, Plá RR, González CM, Carreras HA, Orellana L (2002) Atmospheric quality and distribution of heavy metals in Argentina employing *Tillandsia capillaris* as a biomonitor, *Environ. Pollut.* 120:59–68.

Rodríguez JH, Weller SB, Wannaz ED, Klumpp A, Pignata ML (2011) Air quality biomonitoring in agricultural areas nearby to urban and industrial emission sources in Córdoba province, Argentina, employing the bioindicator *Tillandsia capillaris*. *Ecol. Ind.* 11:1673–1680. doi:10.1016/j.ecolind.2011.04.015

Shafiq M, Iqbal MZ (2003) Effects of automobile pollution on the phenology and periodicity of some roadside plants. *Pak. J. Bot.* 35(5):931–938.

Classification in fMRI studies: in search of brain informative regions.

Juan E. Arco¹, Paloma Diaz-Gutierrez¹, Javier Ramírez² and María Ruz¹

¹ Mind, Brain and Behaviour Research Center (CIMCYC)

² Signal Theory, Networking and Communications
University of Granada
jearco@ugr.es

Abstract. Multi-voxel pattern analysis (MVPA) has been successfully applied to neuroimaging data due to its larger sensitivity compared with univariate traditional techniques. Whereas in clinical applications yielding the highest accuracy as possible is the main aim, in the neuroscientist context it is of great importance to explore how information is distributed in the brain during a certain cognitive process. For this reason, it is necessary to select an approach that leads to a good classification performance and provides information about the brain regions that guides the decision of the classifier. Searchlight analysis is one of the most appealing techniques because it addresses both the localization goal and the coactivation of adjacent voxels in functional Magnetic Resonance Imaging (fMRI) studies. However, the locally multivariate nature of Searchlight can lead to limitations identifying informative regions. An alternative is to employ a Multiple Kernel Learning (MKL) algorithm, which uses a classifier for each previously-defined brain region and combines them to build a global decision function. In this work, we compared the performance of both methods and evaluate their suitability to infer the functional role of brain regions.

Keywords: Searchlight, Multiple Kernel Learning, fMRI, MVPA

1 Introduction

In the last few years, the use of multivariate techniques to neuroimaging data has increased considerably. The study of the patterns distribution instead of the mean activation itself in a certain brain region provides a larger sensitivity than the univariate techniques based on the General Linear Model (GLM, [1]). However, it is necessary to adapt the methods available in machine learning to the requirements of the neuroscientist context: the main purpose is not just to find if two brain patterns are different but the location of these differences. In the present work, we used two approaches (e.g., a Searchlight [2] and a Multiple Kernel Learning) in order to evaluate their ability to find specific regions that allow specific inference.

2

2 Database

The fMRI dataset contained data from 24 subjects while performing a task. The task contained two events in each trial: a word (positive, negative or neutral in valence) and second two numbers, to which participants had to respond, during 8 runs. To test the reliability of the different approaches, we focused on the comparison between the positive vs. negative valence of the words, that were equated in number of letters, frequency of use and arousal.

3 Methods

3.1 Searchlight

We employed a searchlight analysis across the whole brain. We used The Decoding Toolbox (TDT, [3]) to create spherical regions with a radius of 12 mm. This sphere is moved across the whole brain, performing a classification in each position using the voxels inside the sphere. The resulting accuracy is assigned to its central voxel, obtaining an accuracies map. In each position, the classification was performed using binary SVM, with a leave-one-run-out cross-validation used to compute model performance, assessed in terms of balanced accuracies.

3.2 Multiple Kernel Learning

In case of MKL, all modelling steps were performed based on PRoNTo software [4], which relies on an adaptation of the “simpleMKL” version presented in [5]. The brain is divided using functional brain regions defined in [6]. Each region corresponds to a different decision function (e.g. SVM classifier), which are assembled to build a global model. The contribution of each region to the final classification results in a weights map showing how informative a region is.

3.3 Statistical method

We used a non-parametric approach based on permutation testing to assess the statistical significance [7]. For each subject, a searchlight analysis was performed after shuffling the labels of the two classes, obtaining a null accuracy map. We carried out this step 100 times for each subject, yielding 100 permuted accuracy maps. Next, we randomly picked one map for each participant following a Monte-Carlo resampling (with replacement), averaging them to obtain a group permuted map. We then repeated this procedure 50000 times, building a empirical chance distribution for each voxel. The resulting maps were FWE-corrected ($p < 0.05$) for multiple comparisons at a cluster level. All voxels from maps obtained with the real labels that surpass this threshold significance was marked as significant. For MKL, the procedure was essentially the same but using the weights maps instead of the accuracy maps.

4 Results

There is a number of regions where both approaches were able to identify the differences between the two classes. Figure 1 shows the overlapping between these regions, but a large difference between the two methods as well. Figure 2 summarizes the voxels distribution in the overlapped regions, finding larger differences in the Supplementary Motor Area (SMA) and the Superior Frontal Gyrus (SFG). MKL leads to a sparse solution due to the L1-regularization on the weights of each brain region. Besides, regions defined by the atlas are larger than the spheres in the searchlight. It is possible that part of these anatomical regions contain information, but in case the region as a whole is not informative, the MKL will be labelled as a zero-contribution region. Searchlight may find subdivisions in these regions since it does not rely on any assumption regarding the distribution of the brain.

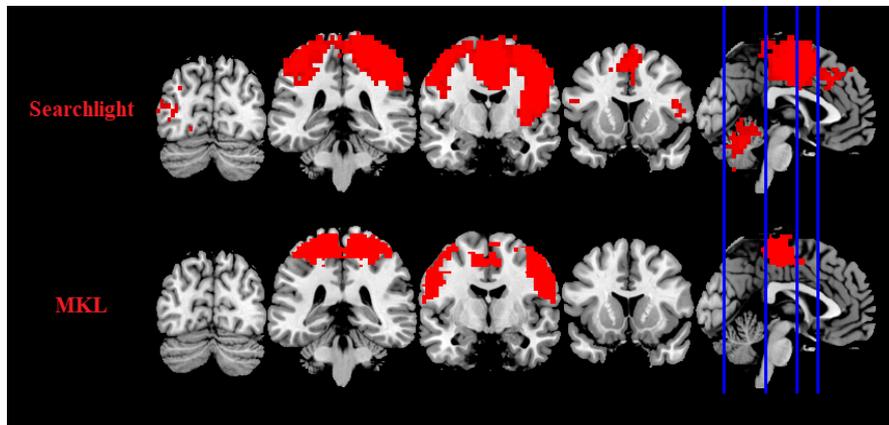


Fig. 1. Significant regions obtained by the Searchlight and the MKL approaches. The lower sensitivity of the MKL is mainly due to the sparse solution that the L1-regularization leads to.

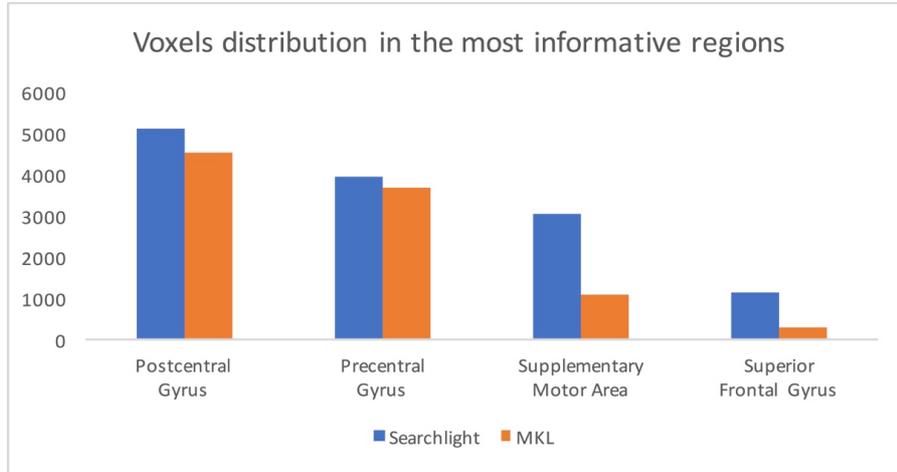


Fig. 2. Voxels distribution in the most informative regions. Differences between the Searchlight and the MKL are larger in the SMA and the SFG, obtaining similar results in the Postcentral and Precentral Gyrus.

5 Conclusion

In this work, we compared searchlight and MKL methods to evaluate the spatial distribution of brain activity patterns in an fMRI study. Searchlight shows a larger sensitivity: since it does not rely on any assumption regarding the distribution of the brain, it can find information in subdivisions of the atlas regions that the MKL consider. Future research should evaluate the MKL approach into an atlas that matches better the functional organization of the brain, i.e. building an atlas from functional data of each subject. Additionally, a different regularization that does not enforce sparsity as L1-normalization does may result in results more similar to those obtained by searchlight.

References

1. Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.-P., Frith, C.D., and Frackowiak, R.S.J. Statistical parametric maps in functional imaging: a general linear approach. *Human Brain Mapping* 2, 189–210 (1995).
2. Kriegeskorte, N., Goebel, R., and Bandettini, P. Information-based functional brain mapping. *Proc Natl Acad Sci U S A* 103(10), 3863–8 (2006).

3. Hebart, M.N., Görgen, K. and Haynes J.D. The Decoding Toolbox (TDT): A versatile software package for multivariate analyses of functional imaging data. *Front. Neuroinform* 8(88), (2015).
4. Schrouff, J., Rosa, M.J., Rondina, J.M., Marquand, J.F., Chu, C., Ashburner, J., Philips, C., Richiardi, J. and Mourão-Miranda, J. PRoNTTo: Pattern Recognition for NeuroImaging Toolbox. *Neuroinformatics* 11, 319-337 (2013).
5. Rakotomamonjy, A., Bach, F.R., Canu, S. and Grandvalet, Y. SimpleMKL. *Journal of Machine Learning Research* 9, 2491-2521 (2008).
6. Urchs, S., Christian, D., Yassine, B. & Pierre, B. Group multiscale functional template generated with BASC on the Cambridge sample. <http://dx.doi.org/10.6084/m9.figshare.1285615> (2015).
7. Stelzer, J., Chen, Y. and Turner R. Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): Random permutations and cluster size control. *NeuroImage* 65, 69-82 (2013).

MOLECULAR MODELING OF THE ADSORPTION OF 5-AMINOSALICYLIC ACID ON THE KAOLINITE AND HALLOYSITE NANOTUBE SURFACES

ANA BORREGO-SÁNCHEZ^{1,2}, MAHMOUD E. AWAD^{1,2,3}, ELISABETH ESCAMILLA-ROA¹, ESPERANZA CARAZO², C. IGNACIO SAINZ-DÍAZ¹

¹Instituto Andaluz de Ciencias de la Tierra (CSIC-University of Granada), Av. de las Palmeras 4, 18100 Armilla, Granada (Spain).

²Department of Pharmacy and Pharmaceutical Technology, Faculty of Pharmacy, University of Granada, Campus de Cartuja s/n, 18071 Granada (Spain).

³ Department of Geology, Faculty of Science, Al Azhar University in Cairo, Nasr City, 11884, Egypt.

1 INTRODUCTION

5-aminosalicylic acid (5-ASA) is an anti-inflammatory drug widely used in the treatment of different diseases, as Crohn's diseases, chronic bowel ulcerative colitis and proctitis [1]. For the treatment of the Crohn's disease and chronic bowel ulcerative colitis disease is administered orally [2] and the 5-ASA is rapidly absorbed in the stomach and in the small intestine. However, the drug adsorption at the level of the large intestine and the colon is very important for the treatment of these diseases [3]. Therefore, it may be necessary the use of excipients to achieve colon targeted drug delivery systems of the 5-ASA drug. For it, kaolinite and halloysite nanotubes was proposed as a good candidates. Halloysite ($\text{Al}_2\text{Si}_2\text{O}_5(\text{OH})_4 \cdot n\text{H}_2\text{O}$) is a multilayer nanotubular clay mineral resulting from the wrapping of 1:1 layers of kaolinite with 500-1000 nm in length and 15-100 nm in inner diameter [4]. Kaolinite and halloysite nanotubes are common excipients in pharmaceutical products and can modify drug bioavailability, due to their ubiquity, low-cost, high surface with activity and biological safety. In addition, they can retain organic molecules and, after administration, release the retained bioactive compounds under controlled conditions [5]. To date several experimental studies have been carried out to study the interaction between 5-ASA and halloysite nanotubes [6-7]. After the experimental studies, it was necessary to carry out theoretical studies that explain the interatomic interactions that occur in the processes of 5-ASA adsorption on the kaolinite and inside the halloysite nanotube, when the 5-ASA is adsorbed inside of the halloysite nanotube [3].

Therefore, the aim of this work was to perform molecular modeling studies by methods based on atomistic force fields for molecular modeling and quantum me-

chanics calculations to predict the interaction between the drug 5-ASA and the kaolinite and halloysite nanotube excipients, being able to explain the use of these excipients as colon targeted drug delivery systems.

2 METHODOLOGY AND MODELS

Two conformer of the molecular structure of 5-ASA molecular structures were considered as neutral forms. The crystal structure of kaolinite was obtained from the experimental neutron powder diffraction and rietveld refinement [9]. The halloysite structure was generated from the atomic coordinates of a slide of a halloysite from a previous work [10]. Periodic boundary conditions were applied to create crystal structures of kaolinite and a cylinder halloysite with an internal diameter of 27 Å, which formation procedure was explained previously elsewhere [10]. Although the internal diameter of a natural halloysite nanotube is around 15-50 nm, our model can be considered a good scenario to reproduce the interactions at molecular level of the adsorption process.

In the kaolinite, a 3x2x1 supercell was generated, composed of 156 atoms distributed within the composition $Al_{24}Si_{24}O_{60}(OH)_{48}$. While, in the halloysite nanotube, a 1x1x2 supercell was generated, $Al_{152}Si_{152}O_{380}(OH)_{304}$, with 1292 atoms to avoid intermolecular interactions between adsorbates of vicinal cells.

Both conformers of 5-ASA (F1 and F2) (Figure 1) were optimized in a periodical box of $30 \times 30 \times 30 \text{ \AA}^3$. Kaolinite and halloysite crystal structures were optimized with the CVFFH and Compass force fields. For non-bonding interactions, the coulomb and van de Waals interactions were calculated by the Ewald method with a cut-off of 12 Å. As well as, these structures were optimized with quantum mechanical calculations by using Density Functional Theory (DFT) with CASTEP and DMOL3 codes, within the generalized gradient approximation (GGA), the Perdew–Burke–Ernzerhof (PBE) correlation exchange functional [8].

The adsorption of both conformers of 5-ASA (F1 and F2) were placed in different positions on the kaolinite and in the internal surface of halloysite, using CVFFH for the adsorption complexes with the kaolinite and Compass FF for the adsorption complexes with the halloysite nanotube, which have provided good results in previous studies [11].

3 RESULTS AND DISCUSSION

Both conformers of the 5-ASA molecule were optimized with CVFFH and Compass FF, finding a similar energy for both of them, being F1 0.46-2.4 kcal/mol more stable than F2, respectively (Figure 1). Similarly, the kaolinite structure was optimized with CVFFH (Figure 2) and the halloysite nanotube structure was optimized with CASTEP (Figure 3).

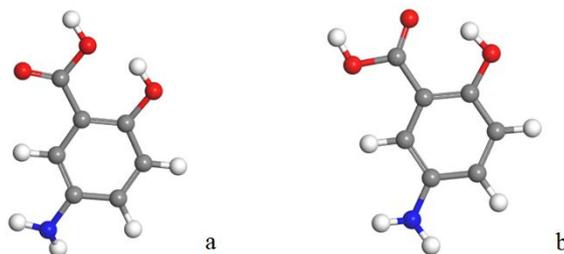


Fig. 1. Molecular structures of F1 (a) and F2 (b) conformers of 5-ASA after its optimization.

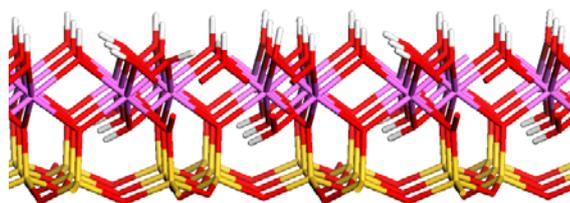


Fig. 2. Crystal structure of kaolinite after its optimization with CVFFH.

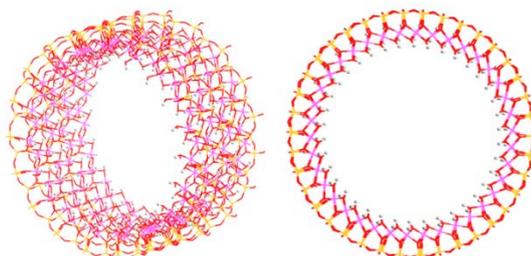


Fig. 3. Crystal structure of 1x1x2 halloysite supercell after its optimization in CASTEP code.

The adsorption between both conformers of 5-ASA and the kaolinite was studied. The adsorption complexes were generated placing the adsorbates in different orientation to determine the most important interaction sites with the mineral. After the optimization (Figure 4), in each adsorption complex, the adsorption energy was studied, as well as, the main interactions between the adsorbates and mineral surface.

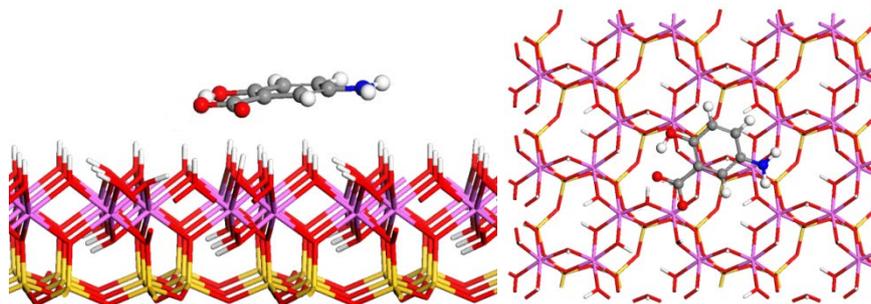


Fig. 4. Adsorption complex of a conformer of 5-ASA on the kaolinite surface after its optimization.

In addition, the adsorption of both structures of 5-ASA with the inside surface of the halloysite nanotube was studied. The adsorbates were placed in different orientations to determine the most important interaction sites with the mineral. After the optimization (Figure 5), of each adsorption complex, the adsorption energy and the main interactions between the adsorbates and mineral surface were studied. The adsorption energy on the kaolinite surface was in the range of 10-14 kcal/mol, whereas on the inner surface of halloysite was in the range of 26-31 kcal/mol depending the orientation of adsorbate.

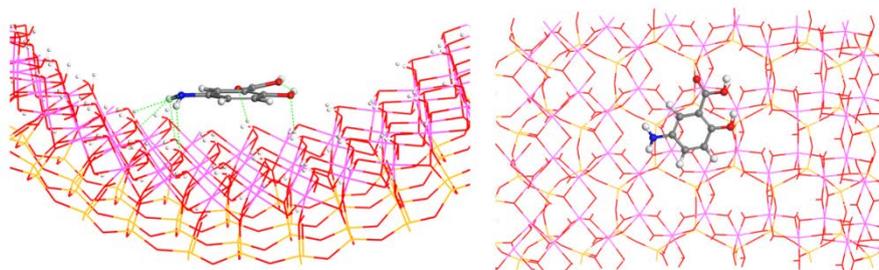


Fig. 5. Adsorption complex of a conformer of 5-ASA in the halloysite nanotube after its optimization.

4 CONCLUSIONS

After the optimization, both conformers yielded negative adsorption energy. This means that the adsorption process is likely to be exothermic and favorable, being the adsorption energy slightly higher for F2 conformer in the kaolinite and in the halloysite. In the kaolinite, the main interactions are hydrogen bonds between the O atoms of carboxylic and hydroxyl groups of 5-ASA and the aluminol H atoms of the mineral surface. In the halloysite, the main adsorbate-surface interactions were strong H bonds between the hydroxyl groups of mineral surface and different groups of the 5-ASA conformers, and electrostatic interactions.

5 REFERENCES

1. Sokollik, C.; Fournier, N.; Rizzuti, D.; Braegger, C.P.; Nydegger, A.; Schibli, S.; Spalinger, J.; Swiss IBD Cohort Study Group. The Use of 5-Aminosalicylic Acid in Children and Adolescents With Inflammatory Bowel Disease. *J. Clin. Gastroenterol.* 2017.
2. Kane, S.V.; Bjorkman, D.J. The efficacy of oral 5-ASAs in the treatment of active ulcerative colitis: a systematic review. *Rev. Gastroenterol. Disord.* 2003, 3, 210-218.
3. Viseras, M.T.; Aguzzi, C.; Cerezo, P.; Viseras, C.; Valenzuela, C. Equilibrium and kinetics of 5-aminosalicylic acid adsorption by halloysite. *Microporous Mesoporous Mater.* 2008, 108, 112-116.
4. Joussein, E.; Petit, S.; Churchman, J.; Theng, B.; Righi, D.; Delvaux, B. Halloysite clay minerals—a review. *Clay Miner* 2005, 40, 383.
5. Viseras, C.; Cerezo, P.; Sánchez, R.; Salcedo, I.; Aguzzi, C. Current challenges in clay minerals for drug delivery. *Appl. Clay Sci.* 2010, 48, 291-295.
6. Aguzzi, C.; Capra, P.; Bonferoni, C.; Cerezo, P.; Salcedo, I.; Sánchez, R.; Caramella, C.; Viseras, C. Chitosan–silicate biocomposites to be used in modified drug release of 5-aminosalicylic acid (5-ASA). *Appl. Clay Sci* 2010, 50, 106-111.
7. Aguzzi, C.; Viseras, C.; Cerezo, P.; Salcedo, I.; Sánchez-Espejo, R.; Valenzuela, C. Release kinetics of 5-aminosalicylic acid from halloysite. *Colloids Surf B Biointerfaces* 2013, 105, 75-80.
8. BIOVIA Materials Studio, version 2016. Dassault Systemes, 2016.
9. Bish, D.L. Rietveld refinement of the kaolinite structure at 1.5 K. *Clays Clay Miner.* 1993, 41, 738-744.
10. Carazo, E.; Borrego-Sánchez, A.; García-Villén, F.; Sánchez-Espejo, R.; Aguzzi, C.; Viseras, C.; Sainz-Díaz, C.I.; Cerezo, P. Assessment of halloysite nanotubes as vehicles of isoniazid. *Colloids Surf B Biointerfaces.* 2017, 160, 337-344.
11. Borrego-Sánchez, A.; Viseras, C.; Aguzzi, C.; Sainz-Díaz, C.I. Molecular and crystal structure of praziquantel. Spectroscopic properties and crystal polymorphism. *Eur J Pharm Sci* 2016, 92, 266-275.

Contrast enhancement methods for images from the light microscope

František Jablončík, Libor Hargaš, Dušan Koniar, Jozef Volák

Department of Mechatronics and Electronics
Faculty of Electrical Engineering, University of Žilina
Žilina, Slovakia
frantisek.jabloncik@fel.uniza.sk

Abstract— This article deals with methods for improving the contrast of the image. The main part of the article is a comparison of methods to improve the histogram properties. The article refers to Homomorphic filtering and its image dividing principle to the part of brightness caused by illumination and a part of brightness caused by reflected light. This algorithm improves the image properties by reducing the dynamic range of the image while increasing the local contrast of the image. Other methods which are mentioned in the article are methods based on adaptive histogram equalization. The article compares the Homomorphic Filtering method and adaptive histogram equalization methods.

Keywords— histogram, contrast enhancement, cilia, equalization, image processing

I. INTRODUCTION

The cilium as kind of epithelium cell is a biological device which serves to defend airways against inspired foreign particles. Cilia are microscopic structures localized on the surface of epithelial cells. They periodically move with synchronous motion with frequency in range 10-30 Hz in vivo. The task of cilia is to move mucus and foreign particles such as bacteria, viruses, dust particles and others against gravitation from lower parts of airways to the oral cavity [1].

The cilium is an elongated microscopic structure with a length of approx. 6-10 μm and width less than 1 μm . Each cell of respiratory epithelium carries on its surface about 200 – 300 cilia [2].

To capture such small microscopic structures as the cilia, it is necessary to provide the best light conditions in the light microscope. However, the resulting image may not be sufficiently sharp, because the sensor is susceptible to the same illumination in high-speed cinematography. For this reason, the resulting image from the sequence may not be sufficiently significant and, for better readability, its properties need to be improved. This is done, for example, by increasing the brightness, contrast, equalization of a histogram or more advanced methods to improve image properties. For further processing, it is needed to highlight the cilia against the background. There exist multiple of methods for enhancement image properties, each of them carries its pros and cons.

II. HOMOMORPHIC FILTERING

A. The mathematical principles of Homomorphic Filtering

The Homomorphic Filtering is based on a model in which the image is represented by two parts in the illumination-reflectance model. This model considers that image consists of two primary components – illumination part and reflectance part. The illumination part of the image is characterized by the amount of source illumination on viewed image scene $i(x,y)$. The reflectance component of the image is a light component of objects reflectance on scene $r(x,y)$. Then the whole image is represented as [3], [4]:

$$f(x,y) = i(x,y) \cdot r(x,y) \quad (1)$$

The main principle of image representation in Homomorphic Filtering model is that intensity of $i(x,y)$ changes slower than $r(x,y)$. Therefore $i(x,y)$ is considered to have more low-frequency components than $r(x,y)$. Homomorphic Filtering aims to reduce the significance of luminance component $i(x,y)$. It is done by reducing the low-frequency components of the image by executing filtering process in a frequency domain. The algorithm of Homomorphic Filtering uses Fourier Transform. Before using the Fourier Transform, it has to be used the logarithm function which changes the multiplication operation of both components of the image to the addition operation [3], [4].

B. The Homomorphic Filtering algorithm order:

1. Natural logarithm of both sides - decouple $i(x,y)$, $r(x,y)$ components of the image

$$z(x,y) = \ln i(x,y) + \ln r(x,y) \quad (2)$$

- Fourier Transform – frequency domain of the image

$$Z(u, v) = F_i(u, v) + F_r(u, v) \quad (3)$$

where $F_i(u, v)$ and $F_r(u, v)$ are the Fourier Transforms of $\ln i(x, y)$ and $\ln r(x, y)$.

- High pass filter of $Z(u, v)$ by means of a filter function $H(u, v)$ in the frequency domain

$$\begin{aligned} s(u, v) &= H(u, v)Z(u, v) \\ s(u, v) &= H(u, v)F_i(u, v) + H(u, v)F_r(u, v) \end{aligned} \quad (4)$$

- Inverse Fourier Transform – filtered image in a spatial domain

$$\begin{aligned} s(x, y) &= \mathfrak{F}^{-1}\{S(u, v)\} \\ s(x, y) &= \mathfrak{F}^{-1}\{H(u, v)F_i(u, v) + H(u, v)F_r(u, v)\} \end{aligned} \quad (5)$$

- Natural exponential function

$$g(x, y) = \exp\{s(x, y)\} \quad (6)$$

- The final filtered enhanced image

$$p(x, y) = g(x, y) \quad (7)$$

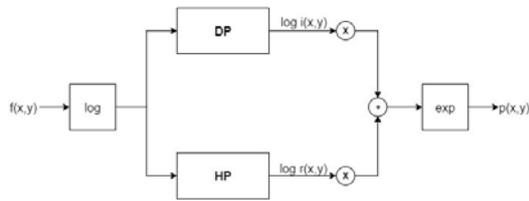


Fig. 1. Algorithm diagram of the Homomorphic Filtering

III. ADAPTIVE HISTOGRAM EQUALIZATION

Adaptive histogram equalization (AHE) is an image processing technique which improves the contrast of the image. The difference between normal equalization and AHE is, that AHE uses several histograms for each corresponding region of the image. AHE in contrast to normal equalization carries the main advantage, where lightness values of images are evenly redistributed by using local histograms. This method is therefore suitable for improving contrast and edges enhancement in all partial region of the image. This method on the other hand also carries disadvantage with a tendency to overamplify noise in relatively homogeneous regions of the image [5].

Contrast limited adaptive histogram equalization (CLAHE) means an improvement to AHE as it limits the amplification by clipping the histogram at a predefined value before computing the cumulative distribution function of pixel values in the neighborhood [6].

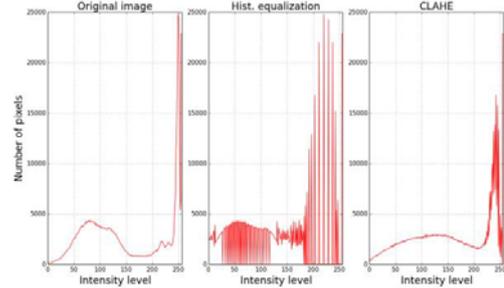


Fig. 2. Comparison of histograms: left- original image histogram, center-histogram of the image after ordinary histogram equalization, right- histogram of the image after CLAHE [7]

IV. COMPARISON OF METHODS OF HISTOGRAM EQUALIZATION

Observing microscopic objects, special live structures like the cilia, requires securing the smallest lighting scattering under the microscope. Under realistic conditions, however, it is not possible to achieve 100 percent homogeneous illumination of objects. It leads to the formation of overexposed parts of image and vice versa to shadowed part. One of the methods to suppress uneven image illumination is Homomorphic Filtering, which principles were mentioned in the previous chapter.

Firstly, it was necessary to collect epithelial samples from the patient's upper airway. After that, it is necessary to observe cilia in the microscope within 5-10 minutes, as they then lose their mobility and gradually dying. To capture fast-moving cilia, it is required a high frame rate (i.e. 240, 500, 1000 frames per second), but the CCD high-speed camera is quite sensitive to sufficient light conditions. To achieve the desired exposure, the camera focuses only on the small area of the image (ROI) at higher frame rates and this is recorded. In our experiment, we recorded video in duration 2-3 seconds. Subsequently, the sequence was divided into individual frames by the computer processing, and individual algorithms for improving the histogram properties were applied to them.

The recorded images were initially colored. Used methods to improve the histogram properties were applied separately to each layer of RGB image. This is more accurate than if the original image had been transformed into a gray-colored image because in that case information from some layers could be lost. Therefore, it is better to perform the method of equalization on individual layers, and if needed, the resulting enhanced RGB image (3 enhanced layers, for each color separately) can be converted to a gray-scale image.

A. Sample Nr.1

As the first sample was chosen an image of respiratory epithelium with cilia. This image was taken under ordinary microscope illumination and its resolution is 450×371 pixels.



Fig. 3. The original image of respiratory epithelium (sample 1)

1) Homomorphic Filtering

Homomorphic Filtering was done for each layer separately and all the operations of the algorithm of this method have been performed, in the order given in Chapter II. First, the image had to be normalized to double values. For sigma value was chosen limits between 0.999 and 1.001, where the transmission function for the Gaussian high-pass filter $H(u, v)$ was calculated at these intervals. The normalized image was divided into the sum of two parts of the image, part of the illumination and part of the reflection by applying the natural logarithm. Afterward, a 2D Fourier Transformation was applied to the image, on which the designed high-pass filter was applied.

Using of Inverse Fourier Transformation lead to change image frequency domain into a spatial domain. As is known, the output of Fourier Transformation, as well as IFT, are complex numbers. For spatial image law, only the absolute size and not the angle of the complex number in the goniometric record is important. Therefore, the complex matrix of the image is expressed only by its absolute value for each pixel.

The last operation of the algorithm was to perform a natural exponent with the previous step. As shown in Fig. 1, each operation of the Homomorphic Filtering algorithm has its own twin that performs exactly the opposite action. Final contrast-enhanced image by Homomorphic Filtering is shown in Fig. 4.



Fig. 4. Homomorphic Filtering applied to the original image (sample 1)

2) AHE method

Another method that improved image properties were AHE. This method divides the image into $M \times N$ areas for which the actual equalization of the histogram is applied. For our experiment, we specified the window size as 8×8 pixels. So, the whole image was divided into the areas (the number of rows divided by $8 \times$ number of columns divided 8) with a custom adaptive histogram equalization [4].



Fig. 5. AHE applied to the original image (sample 1)

3) CLAHE method

The last method in the experiment was AHE with contrast limitation, CLAHE, to prevent noise amplification [5].



Fig. 6. CLAHE method applied to the original image (sample 1)

4) Comparison of methods:

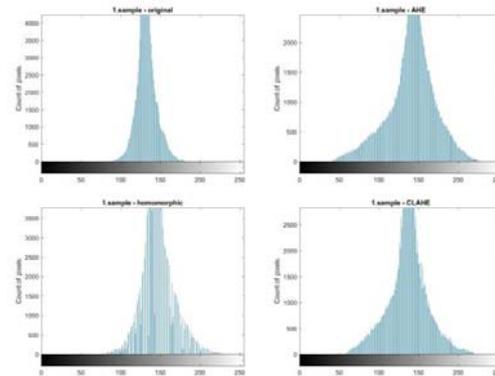


Fig. 7. Comparison of histograms for individual methods of contrast enhancement (sample 1)

As is shown in Fig. 7 histograms of the image after applied methods for enhancement of contrast are different. The widest range of brightness values was achieved at AHE method. The mean value of histogram was best matched to the original image using the Homomorphic Filtering method. As can be seen from the comparison of histograms, the histogram for Homomorphic Filtering has the best corresponding expression of the darker parts of the image. On this basis, it can be said that this method at least highlights undesirable artifacts and image noise.

When viewing the individual final images, a large noise image is noticeable. Contrast limiting adaptive histogram equalization has suppressed this phenomenon. Homomorphic

Filtering has improved the image properties to the point of emphasizing the respiratory epithelium without emphasizing the background of the observed area of the image. The disadvantage is that, in some parts of the image, there is a significant increase in brightness, which degrades other possibilities of image operations.

B. Sample Nr.2

The second sample was a microscopic image of respiratory epithelium with lower resolution (136×160 pixel) and has worse light conditions like at first one. As with the first image here was done the same operations, the resulting differences between histograms are more pronounced.

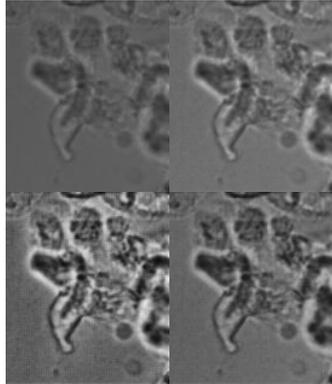


Fig. 8. above left- the Original image of respiratory epithelium (sample 2), above right- Homomorphic Filtering applied to the original image, below left- AHE method applied to the original image, below right – CLAHE method applied to the original image

1) Comparison of methods

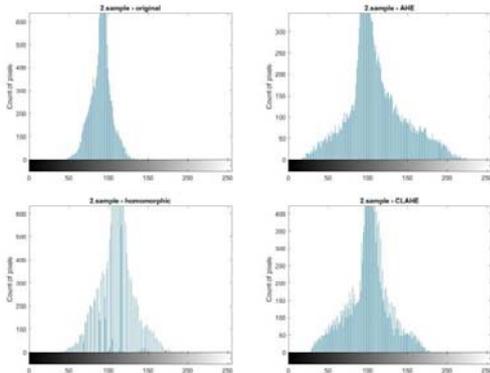


Fig. 9. Comparison of histograms for individual methods of image enhancement (sample 2)

When comparing the individual methods applied to sample number two, the results from the first sample can be confirmed. Unlike the first sample, when the image was well lit, it was captured in worse light conditions. The histogram runs show that the largest range of brightness is the AHE method. The most homogeneous light distribution is the histogram of the relevant Homomorphic Filtering.

V. CONCLUSION

Finally, it can be said that each of these mentioned methods has advantages which are less or more suitable for a given picture. The least appropriate method of contrast enhancement is the AHE method. While the method improved image contrast, it also increased noise in the image. Homomorphic Filtering methods and CLAHE are comparable with their results. By using the Homomorphic Filtering method resulting image becomes brighter than the original image.

Due to low resolution on selected sample images can be decided, that contrast enhancement methods successfully improved image properties. It is seen by eye observation and it is the basis for further processing of medical images in processing and analysis of image by a computer.

To perform further operations from the image, this will subsequently be emphasized, for example by application of an edge detector filter or by an application or by another technique of increasing the sharpness of the image. Afterward, we approach the thresholding, segmentation, and classification of individual objects in the image.

ACKNOWLEDGMENT

Authors of this paper wish to thank for funding from project APVV-15-0462: *Research on sophisticated methods for analyzing the dynamic properties of respiratory epithelium's microscopic elements.*

REFERENCES

- [1] E. Houtmeyers, R. Gosselink, G. Gayan-Ramirez, M. Decramer, „Regulation of motociliary clearance in health and disease,“ *European Respiratory Journal*, pp. 1177-1188, 1999.
- [2] M. A. Chilvers, C. O’Callaghan, [Online]. Available: <http://thorax.bmj.com/cgi/content/full/55/4/314>. [Cit. 11. 9. 2008].
- [3] S.A.M.Saleh, H. Ibrahim: *Mathematical Equations for Homomorphic Filtering in Frequency Domain: A Literature Survey*, IACSIT Press, Singapore, 2012
- [4] R. C. Gonyaley, R. E. Woods: *Digital Image Processing (2nd Edition)*, Pearson, pp. 191.193, 2002
- [5] S. M. Pizer, E. P. Amburn, J. D. Austin, et al.: *Adaptive Histogram Equalization and Its Variations*. *Computer Vision, Graphics, and Image Processing* 39 (1987) pp. 355-368
- [6] N. Kharel, A. Alsadoon, P. W. C. Prasad, A. Elchouemi: *Early diagnosis of breast cancer using contrast limited adaptive histogram equalization (CLAHE) and Morphology methods*. In *Information and Communication Systems, 2017, conference proceedings*.
- [7] https://sleeklens.com/wp-content/uploads/2016/04/histograms_new.png

A Novel Demodulation Technique to Estimate Continuous Blood Pressure from Photoplethysmographic Signals

Fabian Kern¹ and Stefan Bernhard^{1,2}

¹ Pforzheim University,

Tiefenbronnerstr. 65, 75175 Pforzheim, Germany

² Fachbereich Mathematik, Freie Universität Berlin,
Germany

{fabian.kern, stefan.bernhard}@hs-pforzheim.de

<http://www.hs-pforzheim.de>

Abstract. According to the guidelines of the European Society of Hypertension International Protocol revision 2010, the requirements for long time blood pressure measurement are besides simple handling, robustness against movements and accuracy requirements of better than ± 5 mmHg and that the patients motion is not restricted during the measurement. Essentially those requirements are desirable for the reliable interpretation of the blood pressure of hypertensive patients, because the diagnosis often requires long term measurements over a period of 24 hours, where especially non-invasive non-obstructive methods are required to reduce patient load.

The state-of-the-art blood pressure measurement is beside common cuff-based methods the cuff-less estimation based on pulse-transit-time, which is the time a blood pressure wave requires to travel from left ventricle of the heart to another peripheral point in the cardiovascular system. The general relationship between the common pulse-transit-time method and the BP is subject of a series of investigation, whereby different groups of subjects from all ages and health conditions are analyzed in regard to good correlation of pulse-transit-time and systolic blood pressure whereby the correlation coefficient is at least $|r| = 0.85$. However the estimation of the pulse-transit-time has some disadvantages: (i) the dependency on the measuring location and the length of measuring distance, (ii) the choice and method to find minima, maxima and saddle points within the photoplethysmography and electrocardiography signals, (iii) appearance of undetermined fluctuations in the time delay between the pressure wave and the ECG R-peak during blood ejection of the left ventricle and (iv) the influences of auto-regulation on arterial stiffness that produce undetermined drifts of the pulse-transit-time values over time.

Within this work we present a novel method for cuff-less blood pressure measurement by analyzing a single photoplethysmographic signal in the frequency domain and using a frequency demodulation method. The dilatation of the artery and therefore the blood pressure dependent non-linear compliance and non-linear resistance causes non-linear signal deformations. The signal components of the PPG signal with higher am-

plitudes pass the arterial segment faster than those with smaller amplitudes which results in a harmonic phase-shift. Accordingly the harmonic waves are damped by different amounts depending on the dilatation of the artery and hence their amplitudes. As a result the phase-shift and amplitude ratio within each period is dependent on the blood pressure and can be correlated to the invasive blood pressure in a beat-to-beat manner.

To overcome the beat-to-beat method, we analyzed the higher harmonics in the PPG signal using a demodulation technique. Therefore the preprocessing of the PPG signals consists of high-pass filtering with the cut-off frequency $f_G = \frac{1.1}{HR_{max}}$, which produces a signal containing the higher harmonics only. The generated signal thus contains only the non-linear distortions caused by the non-linear pressure dependent dilatation of the arterial segments, leading to a frequency modulation by the arterial pressure, i.e. during systolic blood pressure, the frequency increases, because the artery is dilated under increasing pressure and thus the transit time is reduced. Consequently during diastole the frequency is decreased. The demodulation of the higher harmonics in the PPG signal thus leads to a continuous signal which is influenced by the blood pressure.

Both frequency domain based analysis methods, the phase-shift and the amplitude ratio between the fundamental frequency and the first harmonic are analyzed in a beat-to-beat manner using a dataset from PhysioNet including $n_S = 838$ periods with a total length of 8 mins and 52 sec. The correlation values of the phase-shift analysis obtained from the long-time estimation of the systolic blood pressure were as high as $r = -0.8689$, while the value for the diastolic blood pressure was found to be $r = -0.9040$, while the correlation of the mean blood pressure was $r = -0.9222$. The correlation values of the amplitude ratio of the fundamental frequency and the first harmonic were lower than those of the phase-shift method, obtaining values of $r = 0.6887$ with the systolic blood pressure, $r = 0.6411$ with the diastolic blood pressure and $r = 0.7011$ with the mean blood pressure.

In contrast to the beat-to-beat approaches, the demodulation technique results into a continuous signal, obtaining correlation values to the invasive blood pressure of $r = 0.7578$. The novel continuous demodulation method thus overcomes several shortcomings of the beat-to-beat methods and is expected to improve current methods considerable. According to its simplicity, we assume that the method will improve the measurement conditions and interpretation of blood pressure data of hypertensive patients in near future. The uniqueness of this novel estimation method is that it can be applied to a single PPG signal and that the algorithm only relies on simple and easy to implement signal processing methods to obtain good correlation between the invasive blood pressure and the demodulated PPG signal.

Keywords: blood pressure estimation, cuff-less blood pressure measurement, harmonic pulse wave analysis, long-term blood pressure measurement, non-invasive beat-to-beat blood pressure measurement, photoplethysmography, pulse transit time, pulse wave velocity, continuous blood pressure measurement, demodulation

Bioinformatic analysis of selected aptamer sequences allows the identification of RNA tools for the functional analysis of West Nile virus genomic RNA elements

Cristina Romero-López*, Beatriz Berzal-Herranz and Alfredo Berzal-Herranz*

Instituto de Parasitología y Biomedicina "López-Neyra" (IPBLN-CSIC), PTS Granada, Av del Conocimiento 17, Armilla 18016 Granada, Spain

*To whom correspondence should be addressed: cristina_romero@ipb.csic.es; aberzalh@ipb.csic.es

The WNV is the etiological agent of recurrent outbreaks of febrile illness and encephalitis worldwide. Since the 1999 outbreak in the US, it is considered a global health threat by the WHO. The West Nile Virus (WNV) shares, with other RNA viruses, a compact RNA genome for storing all the required information for the completion of the infectious cycle. For this purpose, RNA viruses have developed an information storage system based on the use of discrete structural units that accomplish well-defined essential functions. This system complements and overlaps the protein coding one and provides an enormous plasticity to the viral genome. These structural/functional units can be conserved among closely related viruses and even among different isolates or strains. Therefore, they are considered potential therapeutic targets for fighting against viral infection.

WNV is an enveloped, single-stranded positive RNA virus belonging to the genus *Flavivirus* (family *Flaviviridae*). It is a mosquito-borne flavivirus that naturally cycles between birds and mosquitoes, although it can infect multiple vertebrate hosts including horses and humans. This genus comprises a large number of viruses, including important human pathogens as Dengue, Japanese encephalitis, Yellow fever or Zika virus, among others.

The WNV genome is a single-stranded RNA molecule of ~11.000 nts, which contains a single open reading frame (ORF) flanked by essential untranslated regions (UTRs). The UTRs are rich in conserved structural RNA elements, which are functionally indispensable for the consecution of critical viral processes, as replication or translation. Such defined RNA structural elements participate in long-distant RNA-RNA interactions that bring close together both UTRs, thus enabling the acquisition of a closed-loop topology required for the achievement of the viral cycle.

The WNV 3'UTR can be subdivided into three autonomously folded regions, domains I-III, with the presence of duplications of structural cassettes (Figure 1). Domain I is located just downstream of the translation stop codon and appears as a hypervariable sequence followed by two conserved stem-loop domains (SL-I and -II) similar in sequence and structure (figure 1). Domain II is moderately conserved and contains a characteristic duplicated structure known as a dumbbell (DB); this is involved in the formation of a pseudoknot (PK) structural element. Domain III is defined by the highly conserved terminal genomic functional elements sHP (short hairpin) and 3'SL. The functions of these structural elements have been studied in depth, and are essential for viral replication and the completion of the viral cycle. However, their role in WNV protein synthesis remains elusive. Therefore, determining their functional implications is a requisite to identify new potential targets for anti-WNV agents. In this context, the use of novel molecular tools based on

nucleic acids, particularly RNA, for elucidating the role of different viral genomic regions has become a successful tactic recently included into the virologists' toolbox.

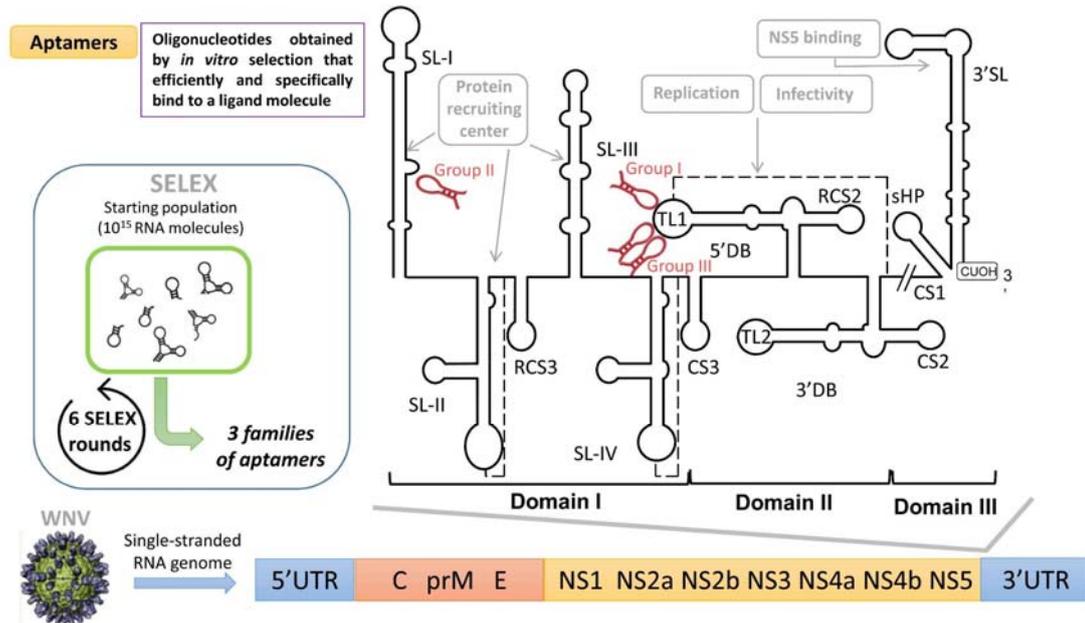


Figure 1. Aptamers are obtained from SELEX (systematic evolution of ligands by exponential enrichment). From a highly heterogeneous pool composed by more than 10^{15} different sequence variants, three groups of molecules (I-III) were defined by bearing common sequence motifs after six rounds of selection. The resulting molecules folded into a single or two stem-loop and targeted highly conserved domains into the WNV 3'UTR.

Aptamers are single-stranded DNA or RNA molecules that bind in a very specific and efficient manner to their targets in a structure-dependent way. Aptamers are isolated by SELEX (Systematic Evolution of Ligands by EXponential enrichment), by which a highly heterogeneous initial population of DNA or RNA, usually composed of 10^{12} to 10^{15} variants, is subjected to iterative selection rounds against a specific target. The SELEX procedure consists in the repetition of a selection cycle composed of binding, partitioning, elution and amplification steps. We have applied a SELEX procedure to identify RNA aptamers against the 3' UTR of the WNV genome by using an initial pool of theoretically more than 10^{15} different molecules, resulting from the randomization of 25 consecutive nucleotides. After six selection rounds, the resulting molecules (aptamers) were subjected to sequence/structure analysis and classification. This bioinformatics study allowed us the identification of three different groups defined by the existence of a common sequence motif, which is complementary to specific sequences within the 3' UTR of the WNV genome (Figure 1). The target regions include the highly conserved 5'DB domain (groups I and III) and the SL-I (group II) (Figure 1). Interestingly, some of the analyzed sequences bear more than one consensus motif, turning these aptamers into interesting multivalent compounds. The pairwise LocARNA algorithm was then used for performing the simultaneous sequence and structural alignment of the selected RNA variants. Two main structural clusters were identified, defined by the acquisition of one or two stem-loops. This result demonstrates the

structural convergence of the RNA pool during the SELEX procedure to target specific and well-defined RNA domains in the 3' UTR. Representative aptamers of the different families and structural groups were further analyzed for their ability to interact with the WNV 3' UTR. *In vitro* binding assays revealed that the selected molecules bound to their target regions with high affinity and efficiency, achieving K_d values in the range of low nanomolar. Finally, the aptamers were characterized for their potential to interfere with viral translation, with the aim of elucidating the role of the targeted domains during WNV protein synthesis. Taken together, our findings highlight the feasibility of using RNA molecules as efficient tools to study viral biology and find new targets for the development of efficient antiviral therapies.

**The effect of c-myc monoclonal antibody for gastric cancer cells in
vivo and in *vitro***

Running title: function of c-myc McAb for gastric cancer

Highlights:

The c-myc McAb prepared in our experiment was high sensitive and specific.

C-myc McAb suppressed the expression of c-myc both in *vivo* and in *vitro*.

C-myc McAb inhibited tumor growth and cell proliferation in a dose-dependent manner.

C-myc McAb induced cell apoptosis in a dose-dependent manner.

Abstract

Purpose: We aimed to investigate the effect of the c-myc monoclonal antibody (McAb) on gastric cancer (GC) cells *in vivo* and *in vitro*.

Methods: C-myc McAb was prepared by immunizing Balb/c mice with c-myc protein. The suspension of SGC-7901 cells was inoculated into nude mice to establish the model with gastric tumor. The anti-tumor rates were calculated and the expression of c-myc was detected by immunohistochemistry in nude mice (*in vivo*). For the SGC-7901 cells (*in vitro*), expression of c-myc was identified by western blot and the cell proliferation was examined by 3-(4, 5-dimethylthiazole-2-yl)-2, 5-diphenyl tetrazolium bromide assay kit. After cell adhesion was calculated by using microscope and cell migration rate was assayed with trans-well inserts, cell cycle and apoptosis were evaluated by flow cytometer.

Results: C-myc McAb was successfully obtained and the subtype was immunoglobulin G1. The expressions of c-myc in *vivo* and *vitro* were decreased

significantly ($p < 0.05$). The tumor growth and cell proliferation were inhibited in a dose-dependent manner. Cell adhesion rate and migration rate were both reduced remarkably ($p < 0.05$) in 2 $\mu\text{g/mL}$ and 4 $\mu\text{g/mL}$ group. At last, the SGC-7901 cells in G_0/G_1 and S phase increased significantly ($p < 0.05$) and apoptosis rates were up-regulated remarkably ($p < 0.05$) in 2 $\mu\text{g/mL}$ and 4 $\mu\text{g/mL}$ group.

Conclusions: The c-myc McAb prepared in our study is available. It can inhibit tumor growth, cell proliferation and induce cell apoptosis. Its effect is associated with down-regulating the c-myc protein level.

Key words: C-myc McAb; Gastric cancer; Proliferation; Apoptosis

Introduction

Gastric cancer (GC) is estimated to be one of the most common and frequent malignant tumor of the digestive system. The incidence and mortality of GC have ranked the second among all tumor diseases worldwide [1-5]. However, it ranks in first place in China [6]. Complete surgical resection remains the cornerstone of GC treatment [7]. But the 5-year survival rates are disappointingly low at 25 to 30% due to locoregional relapse and distant metastases [8]. Therefore, the setting of new targets is necessary for diagnosis, prevention and treatment of GC.

Cellular homologue of avian myelocytomatosis virus oncogene (c-myc) is an important member of the myc family and a master regulator of genes involved in diverse cellular processes in GC [9]. Once c-myc is activated in *vivo* or in *vitro*, it is easy to make the cells far from the normal growth and promote cell malignant transformation to GC [10-12]. Therefore, the inhibition of c-myc plays

a key role for the treatment of GC. Recent experimental data indicate that even a brief inhibition of c-myc expression may be sufficient for stopping tumor growth permanently and inducing regression of tumors [13]. At present, the suppression of c-myc is usually conducted at genetic level including the use of antisense-myc oligonucleotide, c-myc ribozyme and siRNA [14-16]. Few researches are performed to inhibit the c-myc expression on protein level especially by using monoclonal antibody (McAb).

In this study, we prepared the c-myc McAb and measured the influence of it on anti-tumor rate and the c-myc expression in GC model of nude mice (*in vivo*). Additional, we also investigated the effect of c-myc McAb on SGC-7901 cells (*in vitro*) with the indexes of c-myc protein expression, growth rate, adherence rate, migration rate, cell cycle and cell apoptosis. We are committed to reveal the effect of c-myc McAb on GC and provide feasible approaches for the oncotherapy of GC.

Materials and Methods

Preparation of c-myc McAb

All experiments involving animals were approved by the Institutional Animal Care and Use Committee of Renji Hospital Affiliated to Shanghai Jiao Tong University of Medicine. Mice used in this study were obtained from Animal Science Laboratory of Shanghai Jiao Tong University. The c-myc proteins prepared in *E.coli* BL21 were used as immunogens. Female Balb/c mice aging from 6-8 weeks were immunized intraperitoneally with 50 μ g c-myc proteins mixed with equal volume of complete Freund's adjuvant (CFA). The immunization was repeated with the same

amount of immunogens in incomplete Freund's adjuvant (IFA) at 14 d. A final immunization was performed with 100 µg c-myc proteins in IFA at 28 d. When the serum titer (at 35 d measured by enzyme-linked immuno sorbent assay (ELISA), Cistron Biotechnology, Pine Brook, NJ, USA) reached 64000, the booster immunization was conducted with c-myc proteins.

Five days after the boost injection, spleen cells were isolated and fused with SP2/0 myeloma cells. The cells in the period of logarithmic phase were screened for positive clones by ELISA. At last, the positive cloning cells processed by Silica gel H were inoculated intraperitoneally into unsexed Balb/c mice and ascetic fluid was collected to extract McAb. The concentration of McAb was determined by protein assay reagent kit (BCA, Pierce, Rockford, USA).

Characteristic identification of c-myc McAb

Subtype identification: The subtype of purified McAb was determined by Rapid Mouse Ig Isotyping Array 1 according to the manufacturer's instructions (Ray Biotech, Inc., Atlanta, USA).

Antibody titers: Antibody titers were measured by ELISA kit [33]. Each 10 µg c-myc protein was coated and loaded onto polystyrene plates. The plate was washed and blocked for 1 h. Gradient diluted McAb was added and incubated for 2 h. The McAb binding to the antigen was detected with rabbit anti-mouse IgG-HRP (1:5000, Boster, Wuhan, China). Optical density (OD) values were read at 450 nm. The supernatant of medium cultured SP2/0 cells was simultaneously amplified as a negative control and the antibody titers were defined as the maximum dilution of

McAb positive reaction which was defined that the OD value of reaction was twice higher than that of negative control.

Specificity identification: The antigenic specificity of McAb was determined by western blot. Purified c-myc protein was transferred into *E.coli* DH5 α and the transfected cells were lysed in ice-cold radioimmunoprecipitation (RIPA) buffer and centrifuged to collect the supernatant [35]. Protein in supernatant were separated by 16% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and transferred onto nitrocellulose membranes. The membranes were blocked with 5% nonfat dry milk. After several rinses, the membranes were probed with c-myc McAb ($1-5 \times 10^4$ cpm/mL) at 4 °C overnight and incubated with HRP-conjugated goat-anti-mouse IgG (50 μ g/mL, Amersham, Inc.; Cleveland, OH, USA) at 22 °C for 1 h.

The establishment of model for GC in nude mice

A total of forty Balb/e2nu/nu mice aging from four to five-weeks (half male and half female; 18 ± 1.5 g; Animal Science Laboratory of Shanghai Communication University) were purchased for the establishment of GC model. The human GC cell line (SGC-7901cell) in logarithmic phase were trypsinized and resuspended to the concentration of 2×10^7 cells/mL. The cell suspension (0.2 mL) was injected subcutaneously at the root of right leg.

Anti-tumor rates

Mice were divided into 4 groups (10 mice/group, half male and half female) and inoculated intraperitoneally with 0.2 mL c-myc McAb at different concentrations

including low-dose group (10 mg/kg), middle-dose group (20 mg/kg), high-dose group (30 mg/kg) and control group (normal saline instead of c-myc McAb). Mice were killed after 4 times injection weekly by ether exposure. All tumors were excised and weighted to calculate the anti-tumor rates ($\frac{\text{average quality of the control group} - \text{average quality of the treatment group}}{\text{average quality of the control group}} \times 100 \%$).

Immunohistochemistry (IHC)

Paraffin embedded tissues were sliced in 4 μm sections and sections were deparaffinized in xylene followed by graded ethanol and rehydrated in phosphate buffer (PBS, pH 7.5). Then, they were microwaved for 10 minutes for antigen retrieval and blocked for 10 min in normal sheep serum at room temperature. After rinsing in PBS, sections were incubated with c-myc McAb (1: 100) at 4 °C overnight. The next day the sections were rinsed 3 times in PBS and incubated with secondary antibody (goat-anti-mouse biotinylated, 1: 50 in PBS; DAKO, Carpinteria, California, USA) at room temperature for 30min. And then the sections were incubated for another 30 min with HRP-conjugated streptavidin after rinsed 3 times in PBS. The slides were visualized by diaminobenzidine (DAKO, Carpinteria, California, USA) for 5 min and counterstained with hematoxylin for 2 min. Negative controls involved the same procedure by replacing primary antibody with PBS.

Western blot analysis

Gastric carcinoma cell lines SGC-7901 were divided into 3 groups and treated with C-myc McAb for the concentration of 1 $\mu\text{g/mL}$, 2 $\mu\text{g/mL}$ and 4 $\mu\text{g/mL}$, respectively. Each group was divided into 3 parts and cultured for 24 h, 48 h and 72 h

respectively. The human gastric mucosa cell lines HFE-145 were used as blank control. Cells were collected and lysed in ice-cold RIPA. Cell lysates were separated by 16% SDS-PAGE. Separated proteins in the gels were transferred onto nitrocellulose membrane. The membranes were probed with c-myc McAb (1.5×10^4 cpm/mL) at 4 °C overnight and incubated with HRP-conjugated goat-anti-mouse IgG (50 µg/mL, Amersham Health, Buckinghamshire, England) at room temperature for 1 h. The proteins were visualized using an electrochemiluminescence system (Amersham Pharmacia Biotech, Buckinghamshire, UK).

3-(4, 5-dimethylthiazole-2-yl)-2, 5-diphenyl tetrazolium bromide (MTT) assay

HGC-7901 cells and HFE-145 cells were seeded into 96-well plates (10000 cells/well) and cultured at 37 °C in humidified atmosphere of 5% CO₂ after trypsinization. The HGC-7901 cells and HFE-145 cells were both treated with c-myc McAb (1 µg/mL, 2 µg/mL and 4 µg/mL, respectively). Then the medium was discarded and the cells were washed with PBS. After 20 µL MTT (5mg/mL) was added for 4 h, 150 µL dimethyl sulphoxide was used to dissolve formazan crystals. The absorbance was recorded at 570 nm. The growth curves for each dose were plotted by setting OD value as the ordinate and time as the abscissa.

Cell adhesion assay

SGC-7901 cells and HFE-145 cells were digested with trypsin to prepare single-cell suspensions. Single cells were inoculated to 6-well tissue culture plates (5×10^5 cells/well) and incubated at 37° C for 12 h. When cells were grown to approximately 90% confluence, the cells were treated with c-myc McAb (1 µg/mL, 2

$\mu\text{g/mL}$ and $4 \mu\text{g/mL}$, respectively) for 4 h. Cells were counted microscopically after rinsed in PBS 3 times and digested by trypsin. The formula of $\frac{\text{numbers of total adhesion}}{\text{numbers of total cell}} \times 100\%$ was used to calculate the adhesion rates.

Cell migration assay

Cell migration assays were performed using modified Boyden chambers with a polycarbonate membrane-transwell. The lower chamber was filled with 500 μL mouse embryonic fibroblasts. Matrigel (Becton Dickinson Company, Bedford, USA) were mixed with serum-free medium at 37°C for 30 min and then added into the upper chamber. Cells were released with trypsin and resuspended at a final concentration of 5×10^5 cell/mL in serum-free medium. C-myc McAb ($1 \mu\text{g/mL}$, $2 \mu\text{g/mL}$ or $4 \mu\text{g/mL}$) were added to upper chambers and incubated with the cells at the migration period. At last, the upper surface of the membrane was wiped with a cotton-tipped applicator to remove nonmigratory. The migrant cells were tained for 30 min with methyl alcohol and stained with hematoxylin. The number of stained cells was counted with a microscope. Each determination represented the average of three individual wells.

Cell apoptosis and cell cycle assays

Cell apoptotic and cell cycle was analyzed by flow cytometry method (FCM). HGC-7901cells and HFE-145 cells were preconditioned with c-myc McAb for the concentration of $1 \mu\text{g/mL}$, $2 \mu\text{g/mL}$ and $4 \mu\text{g/mL}$ respectively. After digested by trypsin, the cells were washed with PBS and centrifuged to discard supernatant, then, the cells were re-suspended in 1.0 mL hypotonic propidium iodide solution (PI, Sigma-Aldrich, St Louis, USA) gently for 20 min. The PI fluorescence of individual

nuclei was measured by FACSCalibur flow cytometer (BD Biosciences Immunocytometry Systems, San Jose, CA) according to the manufacturer's instructions.

Statistical analysis

Statistical analysis was performed using SPSS11.0 software. All data were expressed as mean \pm SD (standard deviation). T-test and one-way analysis of variance were conducted at each group. $p < 0.05$ was considered as statistically significant.

Results

Production of c-myc McAb

We successfully obtained a hybridoma cell line which could secrete c-myc McAb steadily. The purity of c-myc McAb extracted from mice ascites reached 97 % and the concentration was 2.3 mg/mL after purification (Figure 1). The subtype of c-myc McAb was IgG1 and the antibody titer was 1: 64000 (Figure 2A). Western blot analysis showed that c-myc McAb in *E.coli* DH5 α bacterial lysate could be identified distinctively by c-myc protein (62 kD, Figure 3).

C-myc McAb affects growth of tumors

The anti-tumor rates in middle-dose group ($40.26 \pm 4.25\%$) and high-dose groups ($44.81 \pm 5.74\%$) were significantly higher than that in low-dose groups ($7.79 \pm 0.92\%$, $p < 0.05$, Table 1).

C-myc McAb inhibits the expression of c-myc both in vivo and in vitro

The brown positive signals of IHC were mostly located in the cell nucleus (Figure 4). The amount of positive signals was obviously decreased in middle-dose and high-dose groups compared with that in normal group and control group. Western blot

analysis in SGC-7901 cells showed that compared with the GC-7901 cell control group, the amount of protein expression was evidently decreased in the 2 µg/mL and 4 µg/mL groups in a dose-dependent manner (Figure 5). However, the expression of c-myc protein had no significant changes as time went on.

C-myc McAb influenced the proliferation of SGC-7901 cells

The growth curves showed that c-myc McAb had cytotoxicity against SGC-7901 cells (Figure 6). When the concentration of c-myc McAb reached 2 µg/mL the cell proliferation was inhibited compared with control group. Since c-myc McAb had little effect on the viability of HFE-145 cells.

C-myc McAb suppressed the adhesion and migration of SGC-7901 cells

Result of the adhesion assay in SGC-7901 cells treated with different concentration showed that the adherence rates of 2 µg/mL group (56.32 ± 8.24) and 4 µg/mL group (48.22 ± 7.15) were significantly lower than that in blank control group (92.15 ± 8.75 , $p < 0.05$, Table 2). However, there were no statistical differences between c-myc McAb groups and blank control group in HFE-145 cells.

Transwell assays showed the migration rate of SGC-7901 cells was significantly lower in the 2 µg/mL group (6.12 ± 0.14) and 4 µg/mL group (5.28 ± 0.25) than that in blank control group (32.25 ± 8.25 , $p < 0.05$, Table 2).

C-myc McAb altered cell cycle and promoted apoptosis of SGC-7901 cells

The mean values of cell cycle analysis were shown in Table 3. The percentage of SGC-7901 cells in the G₀/G₁ phase in the 2 µg/mL group ($58.21 \pm 2.35\%$) and 4 µg/mL group ($60.35 \pm 2.42\%$) was significantly different ($p < 0.05$) than that of the

control group ($24.62 \pm 1.65\%$). Similarly, a significant difference ($p < 0.05$) was noted in the percentage of cells in the S phase in the 2 $\mu\text{g/mL}$ group ($24.54 \pm 3.44\%$) and 4 $\mu\text{g/mL}$ group ($20.64 \pm 1.85\%$) vs. the blank control groups ($58.25 \pm 2.85\%$). However, no significant difference ($p > 0.05$) in the percentage of cells in the G₂/M phase was observed in the c-myc McAb groups, relative to the blank control group. The percentage of HFE-145 cells in each phase in c-myc McAb groups showed no significant difference ($p > 0.05$) compared with that in blank control group. Additionally, the apoptosis rates of SGC-7901 cells in the 2 $\mu\text{g/mL}$ group ($28.42 \pm 2.64\%$) and 4 $\mu\text{g/mL}$ group ($35.65 \pm 5.25\%$) was significantly increased ($p < 0.05$) than that of the blank control group ($5.68 \pm 0.45\%$). But there was no significant difference showed between c-myc McAb groups and control group of HFE-145 cells (Table 4).

Discussion

The GC is one of the most malignant tumor in the world, where it shows the substantial morbidity and mortality [5, 39]. Studies have indicated that the c-myc gene has a close relationship with carcinogenesis [41]. Therefore, reducing the expression of c-myc by using the c-myc McAb may be therapeutic for the treatment of GC. In our experiment, we successfully prepared c-myc McAb, which belonged to IgG1 and could identify c-myc protein specifically. C-myc McAb promoted the anti-tumor rate in *vivo* (mice) and inhibited cell proliferation in *vitro* (SGC-7901 cells). It also decreased the c-myc protein levels both in *vivo* and in *vitro*. What's more, the migration, adhesion were reduced remarkably and cell apoptosis of SGC-7901 cells

was promoted under the action of c-myc McAb.

The c-myc McAb was obtained by immunizing mice intraperitoneally with hybridoma (immunized splenocyte fused with myeloma cell). The antibody titer was 1:64000, the same with the c-myc McAb produced by SUN [46] but significantly higher than that of c-myc McAb (IgG1) produced by Santa Cruz Co. Ltd (1:32000) [46], indicating that the c-myc McAb prepared in our experiment got high sensitivity. At the same time c-myc McAb only reacted with c-myc protein, showing that the McAb was highly specifically. Therefore, it's available to get c-myc McAb in our way.

The overexpression of c-myc can induce malignant transformation and tumor formation in gastric carcinoma [47]. It also contributes to cell growth and cell proliferation [48, 49]. Studies show that down-regulated c-myc protein levels or mRNA levels can decrease tumor growth rates in xenograft tumor [16, 50]. This phenomenon also happened in our study, when the c-myc protein levels were inhibited by McAb, the anti-tumor rates increased significantly ($p < 0.05$). Thus, the c-myc McAb may suppress tumor growth by reducing the expression of c-myc protein.

McAb that binds specifically with its antigen (protein) can selectively neutralize the functions of a protein. It has been reported that c-myc McAb can suppress the expression of c-myc *in vitro* [17]. Hence, c-myc McAb may down-regulating the expression of c-myc so as to suppress the proliferation and malignant phenotype of cancer cells effectively. This hypothesis was verified in our study, the protein levels of

c-myc are decreased in *vitro* under the action the c-myc McAb in a dose-dependent manner and cell proliferation was inhibited when the concentration of c-myc McAb reached 2 $\mu\text{g}/\text{mL}$. What's more, the cell adhesion and migration were receded, indicating that the transfer force of tumor cells was decreased. All the results above were significant different from control group when the concentration of c-myc McAb reached 2 $\mu\text{g}/\text{mL}$. In addition, not only does c-myc induce proliferation in cells, but also it regulates tumor cells to undergo apoptosis [41, 51]. It has now been shown that the down-regulated expression of the endogenous c-myc can trigger cell apoptosis [52]. The use of flow cytometric analysis suggested that 2 $\mu\text{g}/\text{mL}$ or 4 $\mu\text{g}/\text{mL}$ c-myc McAb could strongly induce apoptosis of GC cells. Some studies have reported that alterations of c-myc expression enhances the induction of apoptosis in SGC7901 cells [52, 53]. These findings are similar to those results in our study. What' more, cells enter a programmed apoptosis rapidly initiated by myc both before and after the commitment point in late G₁ [54], this condition is consistent with the result in our study that SGC-7901 cells in G₀/G₁ phase are significantly increased for the effect of c-myc McAb. In addition, the reduced cell adhesion inhibited by c-myc McAb also contributes to the cell apoptosis [55]. These findings suggested that c-myc McAb may result in the inhibition of proliferation, adhesion, migration and apoptosis by down-regulated c-myc protein level.

In conclusion, these study findings have indicated that the c-myc McAb prepared in our study is useful and the c-myc McAb might promote efficient tumor cell death in *vivo* and inhibit tumor cell proliferation in *vitro* by down-regulating c-myc protein

levels. The use of c-myc McAb may be a potential way for reducing tumor growth *in vivo* and for the treatment of gastric carcinoma.

References

1. Jemal, A., et al., *Global cancer statistics*. *CA Cancer J Clin*, 2011. 61(2): p. 69-90.
2. Correa, P., *Human gastric carcinogenesis: a multistep and multifactorial process--First American Cancer Society Award Lecture on Cancer Epidemiology and Prevention*. *Cancer Res*, 1992. 52(24): p. 6735-40.
3. Correa, P., *Clinical implications of recent developments in gastric cancer pathology and epidemiology*. *Semin Oncol*, 1985. 12(1): p. 2-10.
4. Blancato, J., et al., *Correlation of amplification and overexpression of the c-myc oncogene in high-grade breast cancer: FISH, in situ hybridisation and immunohistochemical analyses*. *British journal of cancer*, 2004. 90(8): p. 1612-1619.
5. Bertuccio, P., et al., *Recent patterns in gastric cancer: a global overview*. *International journal of cancer*, 2009. 125(3): p. 666-673.
6. Wei Lv, L.C., *<Current status and recent developments in molecular targeted therapy against gastric cancer.pdf>*. 2007.
7. Hensel, F., et al., *Ten-year follow-up of a prospective trial for the targeted therapy of gastric cancer with the human monoclonal antibody PAT-SC1*. *Oncology reports*, 2014. 31(3): p. 1059-1066.
8. De Vita, F., et al., *Perspectives in adjuvant therapy of gastric cancer*. *Oncology*, 2010. 77(Suppl. 1): p. 38-42.
9. Zimonjic, D.B. and N.C. Popescu, *Role of DLC1 tumor suppressor gene and MYC oncogene in pathogenesis of human hepatocellular carcinoma: Potential prospects for combined targeted therapeutics (Review)*. *International journal of oncology*, 2012. 41(2): p. 393-406.
10. Nair, S. and S. Burley, *Structural aspects of interactions within the Myc/Max/Mad network*, in *The Myc/Max/Mad Transcription Factor Network*. 2006, Springer. p. 123-143.
11. Masauzi, N., et al., *A translocation t (8; 14) and c-myc gene rearrangement associated with the histological transformation of B-cell acute lymphocytic leukemia (FAB-L2) into Burkitt's type (FAB-L3) leukemia*. *Leukemia & lymphoma*, 1997. 27(3-4): p. 357-363.
12. Dang, C.V., *< i> MYC</i> on the Path to Cancer*. *Cell*, 2012. 149(1): p. 22-35.
13. Hermeking, H., *The MYC oncogene as a cancer drug target*. *Current cancer drug targets*, 2003. 3(3): p. 163-175.

14. Cheng, J., et al., *Inhibition of cell proliferation in HCC-9204 hepatoma cells by a c-myc specific ribozyme*. *Cancer gene therapy*, 2000. 7(3): p. 407-412.
15. Watson, P.H., R.T. Pon, and R.P. Shiu, *Inhibition of c-myc expression by phosphorothioate antisense oligonucleotide identifies a critical role for c-myc in the growth of human breast cancer*. *Cancer research*, 1991. 51(15): p. 3996-4000.
16. Wang, Y., et al., *Knockdown of c-Myc expression by RNAi inhibits MCF-7 breast tumor cells growth in vitro and in vivo*. *Breast Cancer Res*, 2005. 7(2): p. R220-R228.
17. FAN YueChao, W.F., ZHANG HongTao,etal., *The Biology Significane of Integrin β 1Expression in Invasive pituitary Adenomas*. *Journal of Brain and Nervous Diseases*, 2006. 13(5): p. 359-361.
18. Wickstrom, E.L., et al., *Human promyelocytic leukemia HL-60 cell proliferation and c-myc protein expression are inhibited by an antisense pentadecadeoxynucleotide targeted against c-myc mRNA*. *Proceedings of the National Academy of Sciences*, 1988. 85(4): p. 1028-1032.
19. Schrama, D., R.A. Reisfeld, and J.C. Becker, *Antibody targeted drugs as cancer therapeutics*. *Nature Reviews Drug Discovery*, 2006. 5(2): p. 147-159.
20. Janghorban, M., et al., *Targeting c-MYC by antagonizing PP2A inhibitors in breast cancer*. *Proceedings of the National Academy of Sciences*, 2014. 111(25): p. 9157-9162.
21. Ma, M.-z., et al., *Long non-coding RNA HOTAIR, a c-Myc activated driver of malignancy, negatively regulates miRNA-130a in gallbladder cancer*. *Molecular cancer*, 2014. 13(1): p. 156.
22. Wu, D., et al., *C-myc suppresses microrna-29b to promote tumor aggressiveness and poor outcomes in non-small cell lung cancer by targeting FHIT*. *Oncogene*, 2014.
23. Sheth, A., et al., *Inhibition of human mitochondrial peptide deformylase causes apoptosis in c-myc-overexpressing hematopoietic cancers*. *Cell death & disease*, 2014. 5(3): p. e1152.
24. Yau, T.M., et al., *Vascular Endothelial Growth Factor Receptor Upregulation in Response to Cell-Based Angiogenic Gene Therapy*. *The Annals of Thoracic Surgery*, 2005. 79(6): p. 2056-2063.
25. Matsubara, J., et al., *Impact of insulin-like growth factor type 1 receptor, epidermal growth factor receptor, and HER2 expressions on outcomes of patients with gastric cancer*. *Clinical Cancer Research*, 2008. 14(10): p. 3022-3029.
26. Dupont, J., M. Karas, and D. LeRoith, *The cyclin-dependent kinase inhibitor p21CIP/WAF is a positive regulator of insulin-like growth factor I-induced cell proliferation in MCF-7 human breast cancer cells*. *Journal of Biological Chemistry*, 2003. 278(39): p. 37256-37264.
27. Liebermann, D.A., B. Hoffman, and R.A. Steinman, *Molecular controls of*

- growth arrest and apoptosis: p53-dependent and independent pathways.* *Oncogene*, 1995. 11(1): p. 199-210.
28. Hale, A.J., et al., *Apoptosis: molecular regulation of cell death.* *European Journal of Biochemistry*, 1996. 236(1): p. 1-26.
 29. Li, X., et al., *Relationship between C-myc and Bcl-2 alterations and biological behavior and apoptosis in gastric cancer.* *Xin Xiaohuabingxue Zazhi*, 1997. 5: p. 773-774.
 30. Hemann, M.T., et al., *Evasion of the p53 tumour surveillance network by tumour-derived MYC mutants.* *Nature*, 2005. 436(7052): p. 807-811.
 31. Vita, M. and M. Henriksson. *The Myc oncoprotein as a therapeutic target for human cancer.* in *Seminars in cancer biology*. 2006: Elsevier.
 32. Degawa-Yamauchi, M., et al., *Serum resistin (FIZZ3) protein is increased in obese humans.* *The Journal of Clinical Endocrinology & Metabolism*, 2003. 88(11): p. 5452-5455.
 33. Watanabe, K.-i., T. Nishizawa, and M. Yoshimizu, *Selection of brood stock candidates of barfin flounder using an ELISA system with recombinant protein of barfin flounder nervous necrosis virus.* *Diseases of aquatic organisms*, 2000. 41(3): p. 219-223.
 34. Clark, M., *Immunosorbent assays in plant pathology.* *Annual Review of Phytopathology*, 1981. 19(1): p. 83-106.
 35. Dressing, G.E., et al., *Progesterone Receptor–Cyclin D1 Complexes Induce Cell Cycle–Dependent Transcriptional Programs in Breast Cancer Cells.* *Molecular Endocrinology*, 2014. 28(4): p. 442-457.
 36. Towbin, H., T. Staehelin, and J. Gordon, *Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications.* *Proceedings of the National Academy of Sciences*, 1979. 76(9): p. 4350-4354.
 37. Kireeva, M.L., S.C.-T. Lam, and L.F. Lau, *Adhesion of human umbilical vein endothelial cells to the immediate-early gene product Cyr61 is mediated through integrin $\alpha v \beta 3$.* *Journal of Biological Chemistry*, 1998. 273(5): p. 3090-3096.
 38. Yebra, M., et al., *Requirement of receptor-bound urokinase-type plasminogen activator for integrin $\alpha v \beta 5$ -directed cell migration.* *Journal of Biological Chemistry*, 1996. 271(46): p. 29393-29399.
 39. Bosetti, C., et al., *Trends in cancer mortality in the Americas, 1970–2000.* *Annals of oncology*, 2005. 16(3): p. 489-511.
 40. Culver, K.W. and R. Michael Blaese, *Gene therapy for cancer.* *Trends in Genetics*, 1994. 10(5): p. 174-178.
 41. Ryan, K. and G. Birnie, *Myc oncogenes: the enigmatic family.* *Biochem. J*, 1996. 314: p. 713-721.
 42. Yau, T.M., et al., *Vascular endothelial growth factor receptor upregulation in response to cell-based angiogenic gene therapy.* *The Annals of thoracic surgery*, 2005. 79(6): p. 2056-2063.
 43. Shibuya, M., J. Yokota, and Y. Ueyama, *Amplification and expression of a*

- cellular oncogene (c-myc) in human gastric adenocarcinoma cells. Molecular and cellular biology, 1985. 5(2): p. 414-418.*
44. Moosvi, Z. and K. Rekha, *c-Myc oncogene expression in selected odontogenic cysts and tumors: An immunohistochemical study. Journal of oral and maxillofacial pathology: JOMFP, 2013. 17(1): p. 51.*
 45. Chen, S., et al., *Mechanistic studies for the role of cellular nucleic-acid-binding protein (CNBP) in regulation of c-myc transcription. Biochimica et Biophysica Acta (BBA)-General Subjects, 2013. 1830(10): p. 4769-4777.*
 46. SUN Huihui, W.X., WANG Huiguo, TAN Jianhua, PIAO Jingai, PIAO Jun, MA Biao, LIU Qingping, LI Wenzhe, *Preparation and identification of monoclonal antibodies against C-myc. Immunological Journal, 2011. 27(8): p. 688-691.*
 47. Ninomiya, I., et al., *Expression of c-myc gene product in gastric carcinoma. Oncology, 1991. 48(2): p. 149-153.*
 48. O'Donnell, K.A., et al., *c-Myc-regulated microRNAs modulate E2F1 expression. Nature, 2005. 435(7043): p. 839-843.*
 49. Boxer, L.M. and C.V. Dang, *Translocations involving c-myc and c-myc function. Oncogene, 2001. 20(40): p. 5595-5610.*
 50. Grand, C.L., et al., *The Cationic Porphyrin TMPyP4 Down-Regulates c-MYC and Human Telomerase Reverse Transcriptase Expression and Inhibits Tumor Growth in Vivo 1 This research was supported by grants from the NIH and the Arizona Disease Control Research Commission. 1. Molecular Cancer Therapeutics, 2002. 1(8): p. 565-573.*
 51. Sakakura, C., et al., *The anti-proliferative effect of proliferating cell nuclear antigen-specific antisense oligonucleotides on human gastric cancer cell lines. Surgery today, 1995. 25(2): p. 184-186.*
 52. Xu, Y., et al., *c - myc - dependent hepatoma cell apoptosis results from oxidative stress and not a deficiency of growth factors. Journal of cellular physiology, 1997. 170(2): p. 192-199.*
 53. Chen, J.P., et al., *Molecular therapy with recombinant antisense c - myc adenovirus for human gastric carcinoma cells in vitro and in vivo1, 2. Journal of gastroenterology and hepatology, 2001. 16(1): p. 22-28.*
 54. Evan, G.I., et al., *Induction of apoptosis in fibroblasts by c-myc protein. Cell, 1992. 69(1): p. 119-128.*
 55. Peluso, J.J., *Putative mechanism through which N-cadherin-mediated cell contact maintains calcium homeostasis and thereby prevents ovarian cells from undergoing apoptosis. Biochemical pharmacology, 1997. 54(8): p. 847-853.*

Figure legends

Figure.1 Purification of c-myc McAb; lane 1-Marker; lane 2-C-myc McAb; lane 3-Ascities;

Figure.2 Identification of the subtype of c-myc McAb by using microarrays; (A) Types of protein were showed in the isotyping array map, such as IgG, IgA, IgE, IgM, IgG2b, IgG1, IgG2a, IgG3; (B-D) In order to state the specific of c-myc McAb, condition medium of SP2/0 cell (B), ascetic fluid (C) and c-myc McAb (D) was added;

Figure.3 Expression of c-myc McAb analyzed by western blot; lane 1-Marker; lane 2- Bacterial lysate of DH5 α ; lane 3-Bacterial lysate of DH5 α transfected by c-myc; lane 4-C-myc protein;

Figure.4 The c-myc expression level analyzed by IHC in different dose of McAb (400 \times); (A) The control group was the normal tumor tissue without McAb added; (B) PBS was instead of primary antibody in the normal saline group; (C-E) The primary antibody and secondary antibody were added at low-dose (C), middle-dose (D) and high-dose (E).

Figure.5 The c-myc expression level analyzed by western blot in different dose of McAb at different point of time; The expression of c-myc protein in SGC-7901 cells was measured at 24h (A), 48h (C), 72h (E); β -actin was use for the internal control; lane 1-1 μ g/mL group; lane 2-2 μ g/mL group; lane 3-4 μ g/mL group; lane 4- SGC-7901 cell control group of; lane 5- HFE-145 cell control group.

Figure.6 Cell growth curves of SGC-7901 cells and HFE-145 cells with different concentrations of c-myc McAb; The OD value of (A) SGC-7901 cells and (B) HFE-145 cells was measured per day for 5 d.

Tables and table legends

Table 1. The tumor suppression of c-myc McAb *in vitro* ($x \pm s$, $n = 10$)

	Quality of tumor (g)	anti-tumor rate (%)
control group	1.54 ± 0.11	0
low-dose group	1.42 ± 0.12	7.79 ± 0.92
middle-dose group	0.92 ± 0.31*#	40.26 ± 4.25#
high-dose group	0.85 ± 0.15*#	44.81 ± 5.74#

* $P < 0.05$: compared with the control group; # $P < 0.05$: compared with the low-dose group; $x \pm$

s : the mean ± SD format; n : the numbers of the tumor;

Table 2. The results of cell adhesion rate and cell migration rate (%), $x \pm s$, $n = 3$)

Cell	blank control group		1 µg/mL group		2 µg/mL group		4 µg/mL group	
	adhesion	migration	adhesion	migration	adhesion	migration	adhesion	migration
	rate	rate	rate	rate	rate	rate	rate	rate
SGC-7901	92.15 ±	32.25 ±	87.42 ±	30.56 ±	56.32 ±	6.12 ±	48.22 ±	5.28 ±
	8.75	8.25	8.62	6.54	8.24*	0.14*	7.15*	0.25*
HFE-145	94.28 ±	0.28 ±	89.24 ±	0.35 ±	90.22 ±	0.22 ±	93.46 ±	0.27 ±
	9.42	0.02	6.53	0.35	6.36	0.06	5.37	0.04

* $p < 0.05$: compared with saline group; %: the percentage of the result; $x \pm s$: the mean ±

SD format; n : samples was in triplicate;

Table 3. The distribution of cell cycle (%), $x \pm s$, $n = 3$)

G0/G1	S	G2/M
-------	---	------

SGC-7901	24.62 ± 1.65	58.25 ± 2.85	17.13 ± 1.35
1µg/mL-SGC-7901	30.26 ± 1.8	57.24 ± 2.1	12.50 ± 1.82
2µg/mL-SGC-7901	58.21 ± 2.35*	24.54 ± 3.44*	17.25 ± 1.92
4µg/mL-SGC-7901	60.35 ± 2.42*	20.64 ± 1.85*	19.01 ± 1.35
HFE-145	29.47 ± 3.48	55.36 ± 5.68	15.17 ± 2.14
1µg/mL-HFE-145	27.88 ± 4.35	57.86 ± 5.42	14.26 ± 1.65
2µg/mL-HFE-145	30.47 ± 3.68	54.32 ± 6.21	15.21 ± 2.35
4µg/mL-HFE-145	31.36 ± 4.58	54.62 ± 3.87	14.02 ± 2.37

* $p < 0.05$: compared with the control group; %: the percentage of the result; $x \pm s$: the mean \pm

SD format; n: samples was in triplicate;

Table 4. The results of cell apoptosis rate (%), $x \pm s$, n = 3)

Cell	blank control group	1 µg/mL group	2 µg/mL group	4 µg/mL group
SGC-7901	5.68 ± 0.45	7.66 ± 0.84	28.42 ± 2.64*	35.65 ± 5.25*
HFE-145	4.58 ± 0.12	5.56 ± 0.78	3.28 ± 0.46	4.29 ± 0.32

* $p < 0.05$: compared with the control group; %: the percentage of the result; $x \pm s$: the mean \pm

SD format; n: samples was in triplicate;

CasANN: Targeted Identification of CRISPR Associated Gene Operons from Next Generation Sequences.

Corey M. Hudson^{1*}, Jerilyn A. Timlin², and Kelly P. Williams¹

¹Sandia National Laboratories, Livermore, CA

²Sandia National Laboratories, Albuquerque, NM

*cmhudson@sandia.gov

Keywords: CRISPR, Cas Gene Annotation, Synthetic Biology, NextGen Sequencing

Abstract.

Background and Motivation:

The identification and annotation of CRISPR associated (*cas*) gene operons, currently requires a considerable amount of effort in sequencing, full genome assembly and annotation. In cases where there is limited interest in the full genome, this level computational burden ultimately produces a small analytical result. The complexity and diversity of *cas* operons, however, mean that assembling larger gene neighborhoods provides information not otherwise available when assembling individual genes.

Many of the recent accomplishments in novel synthetic biology tool development have come from the identification of new *cas* gene families, with more suitable operating preferences. The development of Cas9, Cas13a and Cpf1 based gene synthesis tools may simply mark the beginning of the development of a broad suite of useful prokaryotic tools for gene manipulation. Identifying novel gene synthesis toolsets, requires the ability to assemble and annotate these genes. However, the current sequencing environment, in which genes are left as unassembled reads is limiting our ability to determine and describe these new and important members of these gene families.

In early work in the identification of the *cas* genes from raw samples was Zhang *et al.*, 's (2014) studied *cas* genes in metagenomic samples. More recently, Markova *et al.* (2015) and Shmakov *et al.* (2017) studied novel *cas* genes in one of two ways. The first, is by clustering and assigning reads to gene classes (Zhang *et al.* 2014). The second is by building Hidden Markov Models (HMMs) and Position Specific Scoring Matrices (PSSMs) for known *cas* gene families and annotating well-defined and assembled bacterial and archaeal genomes (Markova *et al.*, 2015; Shmakov *et al.*, 2017).

Methods:

We present, **CasANN**, a software pipeline for targeted assembly of *cas* operons from raw next generation sequencing reads.

CasANN Pipeline:

- 1) CasANN works by first preprocessing reads directly (cleaning, merging, removing adapter sequence, and quality control).

- 2) Reads are translated into amino acid sequences using FragGeneScan, based on known gene models. The implementation of FragGeneScan included in this work allows parallel processing, making use of multiple core systems and retains read level information, to allow direct assembly of reads in downstream steps.
- 3) Amino acid translated reads are used to query 24 unique *cas* HMMs and 191 PFAM/TIGRFAM *cas* and *cas*-related gene families (following Zhang *et al.*, 2014). These families have been thresholded using existing gene assignments, to remove false positives. Additionally, 394 unique alignments, using 7903 genes, were used to create PSSMs and generate RPS-BLAST profiles (following Shmakov *et al.*, 2015).
- 4) In addition to these existing HMMs and PSSMs, a list of CRISPR/*cas* vectors was collected from AddGene. This includes 67 unique amino acid sequences, covering Cas9, CasC2C2 (Cas13a) and Cpf1. These amino acid sequences were used to build a blast database, since they have unique significance, both in ignoring vector sequences and targeting near-neighbors of synthetic biology genes.
- 5) These gene models are back-associated with reads in the next generation dataset.
- 6) These reads serve as a starting point for targeted assembly. Targeted assembly involves BLAST search of targets against the full read dataset, followed by iterative SPAdes assembly. This is performed until the reads fail to elongate the partially reconstructed operon.
- 7) The final product of assembly is then annotated using Prodigal, with common annotation models.

Results:

We have applied CasAnn to a variety of synthetic biology studies of mammalian CRISPR experiments, clinical samples and metagenomic datasets. We show that in the synthetic biology studies, full gene synthesis vectors can be reconstructed from the transfected datasets. In the clinical and metagenomic samples, we demonstrate that large *cas* operons can be discovered from mixed samples. This provides a resource for discovery of novel *cas* architectures as well as associated near-neighbors to commonly used synthetic biology tools.

Discussion:

The software and algorithm presented in this work, is an improvement on previous research by Zhang *et al.* (2014), Markova *et al.* (2015) and Shmakov *et al.* (2017). CasANN provides a needed tool for identifying novel *cas* operons in next generation read datasets, without full assembly and annotation of the associated genome, and uses the most recent databases of Cas gene models. Using a variety of optimizations and parallel computing tools, the software can be run in order to produce and annotate *cas* gene neighborhoods. By constructing operons using targeted assembly, this method will be able to pick important *cas* gene architectures out of difficult to assemble samples. The identification of new Cas proteins, with more industrially suitable properties will likely lead to new classes of synthetic tools.

Characterization of the telomeric transcriptome from *Chironomus sp.* Sequence analyses and structural predictions

Mercedes Iriarte¹, José Luis Martínez-Guitart^{2*} and Mercedes de la Fuente^{1*}

Grupo de Biología y Toxicología Ambiental. Facultad de Ciencias. Universidad Nacional de Educación a Distancia, UNED. Senda del Rey 9, 28040 Madrid, Spain

¹Departamento de Ciencias y Técnicas Fisicoquímicas. Facultad de Ciencias, UNED.

²Departamento de Física Matemática y de Fluidos. Facultad de Ciencias, UNED.

*These authors contributed equally to this presentation.

Abstract. Three different sequences, named TsA, TsB and TsC, exist in the telomeres of the insect *Chironomus riparius*. Transcriptional basal activity has been observed in telomeres with TsA and TsB sequences, being the transcription highly hyperactivated upon heat shock in TsA telomeres. On the other hand, neither basal nor stressed activated transcriptions have been observed in telomere with TsC sequence. A comparative analysis, both sequence and putative secondary structure, of two detected non-coding RNA transcripts, *i.e.* transcripts from TsA and TsB, has been carried out. The analysis of the differences/similarities in each defined structure element in correlation with their physiological behavior, enable us to generate hypotheses about structure-function relationships of these non-coding RNAs.

Keywords. non-coding RNA (ncRNA), telomeric RNA, secondary structure prediction, Chironomidae

1. Introduction

Chironomids are a group of diptera midges with high interest as environmental biomarkers. It is well known that some of them have non-canonical telomeres with transcriptional activity under different stress and constitutive conditions [1]. Specifically, in telomeres of *Chironomus riparius* three different sequences have been reported, named TsA, TsB and TsC. While transcriptional basal activity was observed in telomeres with TsA and TsB, being hyperactivated upon heat shock in TsA telomeres, neither basal nor stressed activated transcriptions have been observed in the telomere with TsC sequence [1]. These sequences are situated in tandem repeat and it has been experimentally proved that, *in vivo*, RNA transcripts with different length can be found, coexisting monomeric units with longer form transcripts. It is well known that secondary structures are highly conserved in evolution for most functional RNAs. An informed determination of the most probable secondary elements of these non-coding RNAs may contribute insights concerning the key structural elements of these molecules. Thus, the goals of this study are:

- (a) to analyze the most likely RNA structure for detected telomeric transcripts in *Chironomus riparius*, either in basal or heat shock temperatures, and to determine the most defined secondary structure elements in the complete folding space for each molecule under both temperature conditions;
- (b) to evaluate the differences/similarities in each defined structure element in correlation with their physiological behavior and,
- (c) to generate hypotheses about structure-function relationships of these non-coding RNAs.

2. Methods

The Minimum Free Energy structures (MFE) for our sequences have been determined at 37°C, 35°C (experimental heat shock temperature) and 20°C (basal temperature for *Chironomus*) by free energy minimization using RNAstructure program [2, 3] with default thermodynamic parameters [4]. For comparative purposes, calculations of MFE at 37°C were also performed under BL* thermodynamic model of Andronescu [5]. Partition function calculations, base-pair probability estimations and determination of Maximum Expected Accuracy structure (MEA) with MaxExpect program, has been carried out by means of RNAstructure suite. RNATips webserver (<http://rnatips.org/> [6]) was also used to analyze temperature-induced changes in the secondary structure of the non-coding RNAs. This method calculates the probabilities of nucleotide coupling based on partition functions. A density plot of significantly changing positions along the sequence is compute for each studied RNA, and distinct clusters of strongly temperature-sensitive positions are identified. The server calculates the probabilities of nucleotides to be in a double-stranded conformation at each temperature within the given range (20-39 C) by using RNAfold tool of the ViennaRNA package [7-9]. Clusters of significantly changing positions are then identified by applying the DBSCAN algorithm [10].

3. Result and Discussion

From the results, it may be highlighted that:

- (a) Sequence alignments might to be indicative of TsB as the evolutionarily oldest sequence. There are a pairwise similarities of the 89% between TsB and TsC, and of 88% between TsB and TsA. Differences are not homogeneously distributed along the sequences. Moreover, 50% of the dissimilarities (11 nt in TsA and 9 in TsC) are accumulated in a region of ~30 pb [1] (see Fig. 1). On the other hand, variations in the sequences between TsA and TsB from *C. riparius* seems to give rise to very distinctive structural motives (see Figs. 2 and 3).
- (b) Experimental data point out a more relevant role for non-coding RNA transcript from TsA (CriTER-A) in heat shock response, since its transcription is hyperactivated under these stress conditions. Regarding this, MFE structures show that region 1 in CriTER-A (see Fig. 2) could exhibit significant structural flexibility, which could be relevant for its function. From RNATips results, different behavior of CriTER-A and CriTER-B structure would be expected in relation to temperature. Thus, CriTER-B show changes of base pairing probabilities between 33°C and higher temperature only in the region 2/2', whereas two zones in CriTER-A exhibit sensitive positions in which base pairing probabilities significantly change with temperature elevation. These positions are clustered in region 3/3' (change from 27°C) and in region 1/1'1'' (change from 33°C) of CriTER-A.

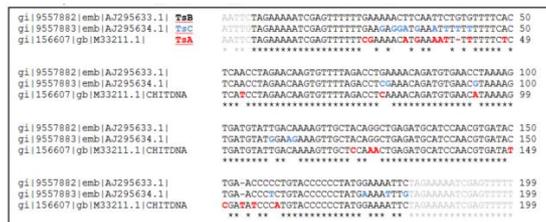


Fig. 1. Three different sequences found in *Chironomus riparius* telomeres, named TsA, TsB and TsC. Telomeres in this Chironomid are characterized by complex DNA repeats of this sequences.

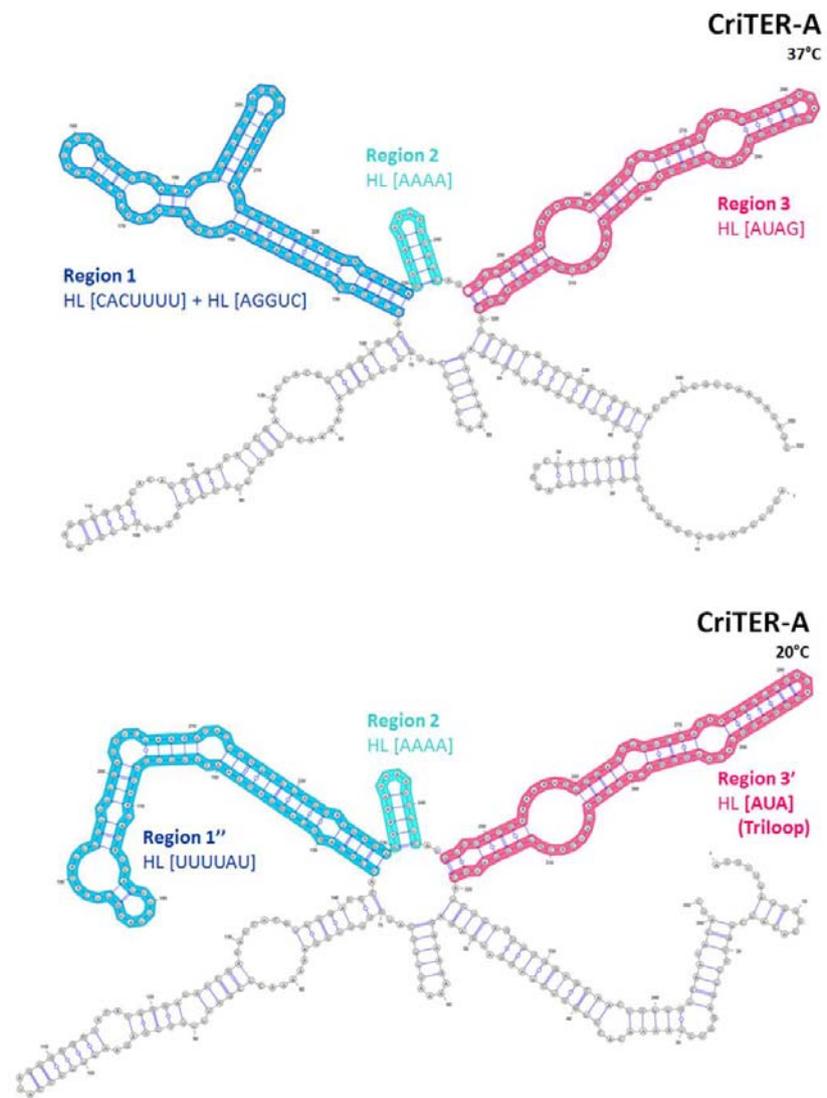


Fig. 2. Minimum Free Energy (MFE) structures predicted at 37°C and 20°C for CriTER-A, non-coding RNA transcripts from *C. riparius* telomere with TsA tandem repeats.

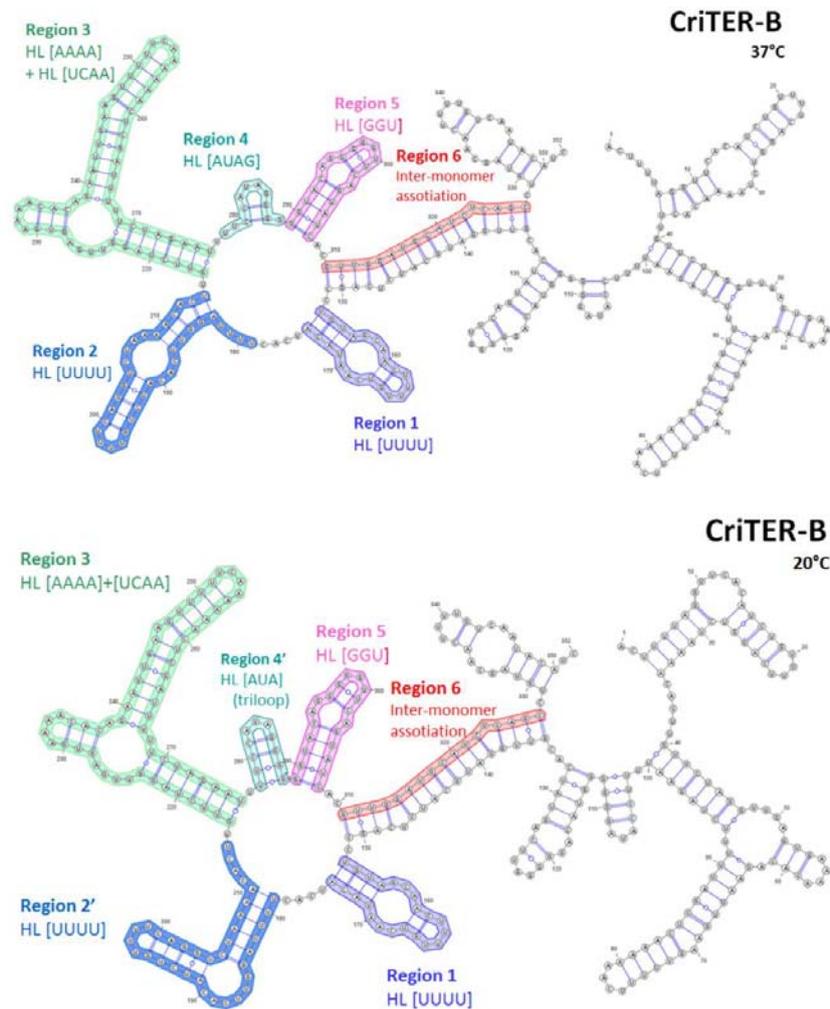


Fig. 3. Minimum Free Energy (MFE) structures predicted at 37°C and 20°C for CriTER-B, non-coding RNA transcripts from *C. riparius* telomere with TsB tandem repeats.

4. Conclusion

As main conclusion, the differential role in heat shock response of CriTER-A non-coding RNA transcribed from TsA sequences of *C. riparius* telomeres could be supported from these results. More specifically, special attention should be paid to the AUA triloop predicted. This triloop [AUA] evokes viral tRNA-like structures. Interestingly, it was early reported that the tRNA-like structure found in BMV and other RNA viruses functions as a telomere, analogously to telomeres of chromosomal DNA [11]. In addition, an AUA triloop in the stem-loop C (SLC) hairpin in the 3'-untranslated region of the plus-RNA strands in brome mosaic virus is the viral polymerase core recognitions site and it seems an essential motif for the production of minus-RNA strand replication intermediates [12]. As well, a short stem-loop structure with the triloop AUA has been related to iron-responsive elements in cellular mRNAs and it has been postulated as a general protein-binding motif [13,14]. Thus, further studies in base to these results are proposed to ascertain the presence of these structural elements and its putative relation to telomere maintenance and heat shock

response. All of this information must be completed by the study and comparison of different telomeric sequences of other Chironomid species to uncover relevant roles of these molecules.

References

1. Martínez-Guitarte, J. L., de la Fuente, M. and Morcillo, G. Telomeric transcriptome from *Chironomus riparius* (Diptera), a species with noncanonical telomeres. *Insect Mol. Biol.*, 23:3, 367–380 (2014).
2. Reuter, J.S., and Mathews, D.H. RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* 11, 129 (2010).
3. Bellaousov, S., Reuter, J.S., Seetin, M.G., and Mathews, D.H.. RNAstructure: Web servers for RNA secondary structure prediction and analysis. *Nucleic Acids Res.* 41, W471–W474 (2013).
4. Turner, D.H., and Mathews, D.H. NNDB: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. *Nucleic Acids Res.* 38, D280–D282 (2010)
5. Andronescu, M., Condon, A., Hoos, H.H., Mathews, D.H., and Murphy, K.P. Computational approaches for RNA energy parameter estimation. *RNA* 16, 2304–2318 (2010).
6. Chursov, A., Kopetzky, S.J., Leshchiner, I., Kondofersky, I., Theis, F.J., Frishman, D., and Shneider, A. Specific temperature-induced perturbations of secondary mRNA structures are associated with the cold-adapted temperature-sensitive phenotype of influenza A virus. *RNA Biol.* 9, 1266–1274 (2012).
7. Gruber, A.R., Lorenz, R., Bernhart, S.H., Neubock, R., and Hofacker, I.L. The Vienna RNA Websuite. *Nucleic Acids Res.* 36, W70–W74 (2008).
8. Hofacker, I.L., Fontana, W., Stadler, P.F., Bonhoeffer, L.S., Tacker, M., and Schuster, P. Fast folding and comparison of RNA secondary structures. *Monatshefte Für Chem. Chem. Mon.* 125, 167–188 (1994).
9. Lorenz, R., Bernhart, S.H., Siederdisen, C.H. zu, Tafer, H., Flamm, C., Stadler, P.F., and Hofacker, I.L. ViennaRNA Package 2.0. *Algorithms Mol. Biol.* 6, 26 (2011).
10. Ester, M., Kriegel, H., S, J., and Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. *AAAI Press*, pp. 226–231(1996).
11. Rao, A.L., Dreher, T.W., Marsh, L.E., and Hall, T.C. Telomeric function of the tRNA-like structure of brome mosaic virus RNA. *Proc. Natl. Acad. Sci.* 86, 5335–5339 (1989).
12. Skov, J., Gaudin, M., Podbevšek, P., Olsthoorn, R.C.L., and Petersen, M. The subgenomic promoter of brome mosaic virus folds into a stem-loop structure capped by a pseudo-triloop that is structurally similar to the triloop of the genomic promoter. *RNA* 18, 992–1000 (2012).
13. Hull, R. *Plant Virology* (Academic Press) (2013).
14. Joost Haasnoot, P.C., Olsthoorn, R.C.L., and Bol, J.F. The Brome mosaic virus subgenomic promoter hairpin is structurally similar to the iron-responsive element and functionally equivalent to the minus-strand core promoter stem-loop. *C. RNA* 8, 110–122 (2002).

Funding

This work was supported by Ministerio de Economía y Competitividad, Spain [Project CTM2015-64913-R from the Ciencias y Tecnologías Medioambientales programme].

Improving the performance of pathway extraction methods by infeasibilities removal

José F. Hidalgo, Francisco Guil, and José M. García

Grupo de Arquitectura y Computación Paralela, Universidad de Murcia, Spain
{jhidalgo,fguil,jmgarcia}@um.es
<http://www.um.es/gacop>

Abstract. Motivation: Genome-scale metabolic network (GSMN) analysis requires efficient calculation of modes and specifically elementary flux modes (EFMs). State of the art methods for retrieving EFMs and pathways for biological studies use a linear programming formulation which is usually solved using the *simplex* method. One of the main drawbacks is the existence of large amounts of infeasible solutions, being an important fraction of them hard-coded in the metabolic reconstruction itself.

Results: Here we propose a method to avoid many of the possible infeasible solutions that can appear extracting modes. Furthermore, the computational cost that can be saved by applying this method can easily be estimated. Results over different case studies found in the literature are provided within the paper.

Availability: We propose a methodology rather than an implementation, and therefore no software code is provided.

1 Introduction

There is no doubt about the importance of achieving a detailed understanding of the functional mechanisms of cellular metabolism [6]. The complete sequencing of the genomes of numerous organisms and the increasing efforts to reconstruct their complete metabolism [20,15,4] are permitting more accurate modeling of metabolic pathways and their application in medicine and biology. Many types of cancer and other diseases are closely related to cell metabolism. A paradigmatic example is the paradox stated as *Warburg effect* [22,21,9]. Warburg discovered how the cells with cancer produce ATP choosing the path of least efficient production [21].

Many strategies have been proposed to analyze cellular metabolism. The available stoichiometric information for the cells of one organism let us build its genome-scale metabolic network (GSMN) where all the relations between enzymatic reactions and the involved metabolites are expressed. Those GSMNs have to be analyzed by extracting subsets of reactions and their corresponding metabolites to be studied one by one. Each of those subsets, named pathways, can be visualized as a partial GSMN with an affordable size. Many authors have

proposed different kind of pathways, being elementary flux modes (EFM) [18,19] the most popular in the recent years. An EFM is a special type of metabolic pathway comprising a minimal number of active reactions that meets the mass balance equation at steady state and thermodynamic-based irreversibility constraints, and cannot be decomposed into smaller solutions. EFMs are an effort to translate a complex network into a canonical expression of vector generators of a solution space. EFMs have been proved to be a useful tool to understand metabolic networks, to model, to simulate and to predict metabolic behaviors within the cells of the live organisms.

In spite of the existence of many successful approaches to extract pathways and specifically EFMs, the increasing size of GSMNs constitutes a challenge to efficiently carry out this task and obtaining significant biological information [11,10]. Many efforts have been made to speed up the extraction of EFMs, but also biological significance of the extracted EFMs has to be improved. Indeed, to keep the focus in the biological significance, the extraction of pathways should avoid to waste computation resources where there are no biological answers. EFMs have also some mathematical properties that allow us to extract them using very well-known techniques like linear programming [13], mixed-integer linear programming [2] or graph theory [1,3].

Linear programming (LP) is on the basis of most extraction methods and mainly for those on the top of the state of the art [14,8]. It must be observed that LP computation is very expensive in floating-point calculations so the strategy has to be balanced between saving LPs and get a number of solutions with a required quality. For that reason, the rate of LP executions per unique feasible solution extracted is one of the most used metrics to measure the efficiency of LP extraction approaches. Most strategies for pathways and EFMs extraction are designed in simple and clear terms but in practice they have to include additional features to circumvent the highly frequent problem of unsolvability observed during the execution. An unsolvable LP problem is termed infeasible in simplex nomenclature. When an LP is posed and it is unsolvable, it is due to the existence of incompatible constraints. Any method based on LP has to avoid to pose infeasible linear programs, but avoiding to produce the infeasibility could be not enough if the original stoichiometric matrices already contain any built-in infeasibility. The origin of those infeasibilities is beyond the scope of this article and we will just propose a method for identifying and removing them.

Our new proposal consists of a method that finds and removes native infeasibilities coming from the metabolic reconstruction. Our proposal can be combined with any other pre-existent strategy by using it as a preliminary step. There are other similar efforts for the constraint-based reconstruction and analysis (COBRA) framework and the elimination of thermodynamically infeasible loops [16].

The paper is structured as follows. First, in Methods section, some terms related to linear programming are revised. Key concepts like additional constraints are overviewed. Discussion introduces our proposal about the elimination of the infeasibility issue to guarantee the feasibility of any linear program. Finally, we estimate the saved efforts by the adoption of our proposal as previous step of

any other strategy to extract pathways. This is shown by studying the impact in some metabolic reconstructions.

2 Methods

2.1 Primary constraints

Constraint based modeling (CBM) uses a stoichiometric matrix S with coefficients for a set of compounds C exchanged in metabolic reactions R . Be \mathbf{v} a vector of flux rates that represents a pathway, v_r is the flux rate for the reaction $r \in R$.

A pathway meets the steady-state condition (Equation 1) when internal metabolites c are balanced and concentration remains constant.

$$\sum_{r=1}^R S_{cr}v_r = 0, \forall c \in C \quad (1)$$

In a pathway, each irreversible reaction only participates with a positive rate (Equation 2).

$$\forall r \in R, v_r \geq 0 \quad (2)$$

Any pathway for the specific GSMN can be represented by a flux rate \mathbf{v} that meets Equations 1 and 2. We refer as *primary constraints* to the two conditions expressed in those two equations. The primary constraints let us construct a linear program that can be solved using the Simplex Algorithm. There are some efficient implementations of simplex algorithm in tools such as CPLEX [7] or CLP-COIN [5].

A pathway is also an EFM if it is non-decomposable, that is, \mathbf{v} is not a positive lineal combination of other minimal flux rate vectors. LP approaches are not only focused on the the extraction of EFMs but also in pathways which do not comply with the non-decomposability condition.

2.2 Linear programming

Given the primary constraints, a clean linear program can be directly posed as shown in Equation 3.

$$\begin{aligned} &\text{Minimize} && \sum_{i=1}^n v_i && (3) \\ &\text{subject to} && S \cdot \mathbf{v} = \mathbf{0} \\ &&& v_i \geq 0 && \forall r_i \in R \end{aligned}$$

This *clean* linear program is minimally restricted by the primary constraints and, therefore, the system $S \cdot \vec{v} = 0$ is homogeneous and usually has infinite solutions (so, in particular the linear programming problem is feasible) but the optimal solution of the linear program is always the trivial one, $\vec{v} = 0$. To get non trivial solutions we must define additional constraints. An additional constraint is a set of individual conditions imposed over the variables of the program.

The main objective of any approach is to find non-trivial feasible solutions of the linear program, which is a sine qua non condition to be a pathway. Metaheuristics are very commonly used to generate new sets of constraints using previous ones and statistics of failure or success during the exploration. Often, LP approaches use previous obtained solutions to evolve them to a set of new constraints, and they use trial and error to fix the potential infeasibilities than can be inferred from the information obtained during the procedure. These strategies need to collect statistics to direct correctly the evolution of the constraints and to prevent signs of exhaustion producing seeds for future iterations. Given the kind and the size of the problem, the collection of additional information along the exploration can produce extra memory consumption and some others inefficiencies.

2.3 Additional constraints

Methods based on linear programming require the execution of the LP solver one time to obtain each pathway. In each execution, the linear program has to be additionally restricted to obtain a new pathway. A constraint is a subset $A \subseteq R$ of reactions that will be forced to be present or absent in the pathway to be obtained. Equation 4 shows an additionally restricted LP in which all the reactions of A have been forced to have a flux rate greater than 0.

$$\begin{aligned} \text{Minimize} \quad & \sum_{i=1}^n v_i & (4) \\ \text{subject to} \quad & S \cdot \mathbf{v} = \mathbf{0} \\ & v_i \geq 0 & \forall r_i \in R \\ & v_j > 0 & \forall r_j \in A, A \subseteq R \end{aligned}$$

The objective of the inclusion of each additional constraint is to induce non trivial, thermodynamically feasible and different solutions from the previous ones. We know that the *clean* linear program (3) has feasible solutions, but adding constraints can make the full linear program (4) infeasible. Another very frequent phenomenon has to do with repeated solutions associated with different sets of constraints (that is, we can impose different constraints to our clean linear program and obtain the same pathway). Different extraction methods of pathways based on linear programming differ precisely on how to generate new additional constraints for the original clean program.

Putting aside the problem of unwanted repetitions, the main objective is to build valid additional constraints for posing feasible linear programs to the LP solvers. We call *feasible constraint* to those that let pose feasible LP.

In order to clarify our approach, we call *positive constraint* to that which forces the presence of one or more reactions in the resulting pathway. Equation 5 shows a minimal additional positive constraint composed only by the reaction $r_s \in R$.

$$v_s > 0 \quad (5)$$

We call *negative constraint* to that which forces the absence of one or more reactions in the final solution. Equation 6 shows a minimal additional negative constraint composed only by the reaction $r_t \in R$. The combination of both types of constraints will be called a *non-positive constraint*.

$$v_t = 0, r_t \in R \quad (6)$$

3 Discussion

3.1 About positive conditions

In order to know if a linear program is feasible or not, the LP solver has to be invoked. The amount of possible additional constraints is extremely high in GSMNs, so passing through all this combinations to classify them as feasible or not is a problem with combinatorial complexity order (Equation 7).

$$O(|R|) = 2^{|R|} - 1 \quad (7)$$

On the other hand, it would be desirable to know when an additional constraint A is incompatible with the other constraints (i.e., produces an infeasible result) to discard it before posing infeasible linear programs, thus saving time and computational resources.

Be $s \in R$ a reaction, s is an *infeasible reaction* if being the only additional positive constraint it produces an infeasible linear program. The Equation 8 shows the LP that has to be solved for each individual reaction.

$$\begin{aligned} \text{Minimize} \quad & \sum_{i=1}^n v_i & (8) \\ \text{subject to} \quad & S \cdot \mathbf{v} = \mathbf{0} \\ & v_i \geq 0 \quad \forall r_i \in R \\ & v_s > 0 \quad r_s \in R \end{aligned}$$

An infeasible reaction makes the problem infeasible either alone or within an additional positive constraint. Also, we have empirically found for many

metabolic reconstructions that no additional positive constraint produces a feasible linear program when an *infeasible constraint* is present. The combination of both statements must be studied.

To know if a reaction is infeasible we have to solve one linear program for every reaction as individual positive constraint. An LP solver reveals if the problem is feasible or not and a solution in the positive case. Be *LPfeasible* a function that using the LP shown in the Equation 8 returns if the LP problem is feasible or not. Algorithm 1 shows how to calculate the set I of infeasible reactions.

Algorithm 1: Calculation of infeasible reactions

Data : Matrix S , set of reactions R

Result: Set I of infeasible reactions

```

 $I \leftarrow []$ ;
for  $s \in R$  do
  if notLPfeasible( $S, s$ ) then
     $I \leftarrow I + [s]$ ;
  end
end

```

Let us use the term *feasible reactions* to refer to those that are not infeasible. R' is the set of feasible reactions from R , calculated like shows the Equation 9.

$$R' = R - I \quad (9)$$

Additional positive constraints have some properties that allow to characterize them as feasible or not without having to invoke to the LP solver for any one of them. Be \mathbf{u} and \mathbf{v} two flux vectors obtained as a solutions of a linear program conditioned with (1), (2) and their respective positive set of constraints A_1 and A_2 . Be $A = A_1 \cup A_2$ another positive condition. The linear program additionally constrained by A has at least as feasible solutions the set of all linear combination of \mathbf{u} and \mathbf{v} with (strictly) positive coefficients. That can be demonstrated in basis to those two feasible LP are homogeneous systems of linear equations (10).

$$S\Delta(\lambda_1\Delta\mathbf{u} + \lambda_2\Delta\mathbf{v}) = \mathbf{0}, \lambda_1, \lambda_2 > 0 \quad (10)$$

$$v_j > 0, v_j \in A$$

The same idea is shown in Figure 1. Given two constraints used respectively as a positive constraint to obtain the respective feasible subgraphs, the resulting graph of the addition of both subgraphs fulfills all the constraints at the same

time. From a graphical perspective, all the additional positive constraints mean the obligation to include each reaction in the final graph or pathway.

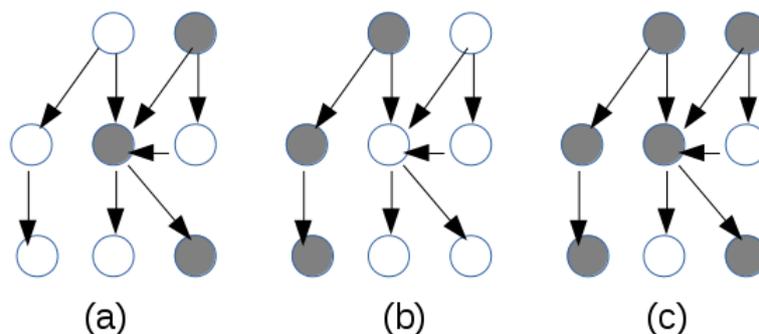


Fig. 1. (a) and (b) represents respectively the graph equivalent to the flux vectors \mathbf{u} and \mathbf{v} . (c) is also a feasible and steady-state subgraph of the GSMN and is the result of the addition of (a) and (b)

As stated above, the inclusion of any infeasible reaction in any additional positive constraint produces an infeasible problem as well. On the contrary, the above discussion tells us that an additional positive constraint is feasible if only includes feasible reactions, that is, $A \subseteq I$.

Given $|I|$ the cardinality of I , the amount of infeasible combinations of reactions is expressed in Equation 11. This (possibly huge) amount of combinations could be automatically discarded as additional constraint from any pathway extraction approach because none of them produces a feasible linear program.

$$2^{|I|} \cdot 2^{|R'|} - 1 \quad (11)$$

Using R' instead of R means also that we are working only with the portion of the GSMN that can produce feasible LP programs. In terms of the model, it can be considered that the data have been curated. This has evident implications on how much biological significance we can obtain using LP even considering R and the full GSMN. Hypothetical GSMN builtin infeasibilities bound the amount, quality and biological significance of the potential solutions that can be found using linear programming.

Finally, it has to be remarked that the combinatorial problem that has been placed at the beginning of the subsection becomes in one of linear complexity due to the fact that it is needed less than $|R|$ LP iterations to classify every reaction as feasible or not. The exploration of any subset of additional constraints reduces drastically the complexity as shown in Equation 12.

$$O(|R|) = \frac{|R|}{3} + 2^{|R'|} - 1 \quad (12)$$

It can be concluded that the important task of classifying each reaction as feasible or not should be done as a first step before trying to find pathways.

3.2 Reversible reactions

Usually, as a previous step to pose the linear program, the stoichiometric matrix is expanded decoupling reversible reactions [3]. The pair of reactions resulting from a reversible reaction has to be considered thermodynamically feasible when they both have positive flux rates. Obviously, the presence of both reactions in the final pathway may have biological sense, but it can be mathematically simplified with the presence of the reaction with the greatest flux rate but subtracting the other rate.

Without loss of generality, let us consider S a matrix already decoupled and R the set of reactions from that S . Be $s, t \in R$ reverse reactions resulting from decoupling a reversible reaction. Analyzing the feasibility of s , an additional positive constraint containing s is feasible if, as well as Equations 1, 2 and 5, there is a feasible solution meeting Equation (13).

$$v_t = 0 \quad (13)$$

As stated above, we are interested in classifying each individual reaction as feasible or infeasible. When the reaction is one of those decoupled reactions, the additional constraint (Equation 13) lets us check if the reaction s is feasible. The same must be checked from the point of view of t .

The core of the function *LPfeasible* must be extended to consider the reversible reactions (see Equation 14).

$$\begin{aligned} & \text{Minimize} && \sum_{i=1}^n v_i && (14) \\ & \text{subject to} && S \cdot \mathbf{v} = \mathbf{0} \\ & && v_i \geq 0 && \forall r_i \in R \\ & && v_s > 0 && r_s \in R \\ & && v_t = 0 && r_t \in R, \quad r_t = -r_s \end{aligned}$$

Regarding the importance of studying separately the reversible reactions, a systematic exploration through the stoichiometric matrix should be done because some reactions can be respectively reverse one of the other despite of they are not strictly the both senses of a unique reversible reaction.

3.3 Estimated impact

The benefits of our proposal can be calculated theoretically as the probability of choosing k reactions from R expecting everyone to be feasible. As shown in Equation 15, the probability depends on $|I|$ and $|R|$ but also on the size k of the set of additional positive constraints A .

$$\frac{\binom{|R'|}{k}}{\binom{|R|}{k}} = \prod_{i=1}^k \frac{|R'| - i - 1}{|R| - i - 1} \quad (15)$$

The lower is the probability, the higher is the savings using R' instead R . The assumption is that the feasible sets of reactions are distributed evenly between infeasible. In general, the distribution of the subsets depends on the strategy to build the additional constraints. The most equitable strategy is to chose reactions randomly.

Equation 16 shows an approximation of the probability for high values of $|R|$ and reasonable values of k . This approximation allows us to observe two facts. The first is the weight of the ratio $|R'|/|R|$ in the formula of probability. The second is that as k rises, the probability of A to be feasible decreases dramatically. In the opposite, knowing the composition of I the probability of build an A feasible is 1.

$$\prod_{i=1}^k \frac{|R'| - i - 1}{|R| - i - 1} \leq \left(\frac{|R'|}{|R|} \right)^k \quad (16)$$

4 Case studies

LP infeasibilities can be produced while the LP is formed but some of them are inherent to the cell reconstruction model. It is evident that in that case computation resources and time are wasted. The benefits of our proposal can be quantified for any stoichiometric matrix.

To measure the real impact of our proposal, the presence of infeasibilities has been checked for some of the most used metabolic reconstructions. Table 1 shows the results of the analysis of the infeasibilities.

Reconstruction	$ R $	$ I $	$ R' $
iAF1260	3234	708	2536
E. coli core	154	16	138
[17]	33	6	27

Table 1. Reactions on each reconstruction. Columns are amount of reactions. Infeasibilities due to unique reactions.

iAF1260 is the reconstruction of the *E. coli K-12 MG1655* organism [4]. *iAF1260* stoichiometric matrix, once decoupled, has 3234 reactions and, *a priori*, any of them is a valid candidate for being included in a pathway that meets the aforementioned primary constraints. After the analysis using the Algorithm 1, 708 reactions have been declared infeasible, that is, the inclusion of any of them (or a combination of them) in a additional constraint for the clean linear program gives an infeasibility as a result.

The *core E. coli metabolic model* [12] is a subset of the genome-scale metabolic reconstruction *iAF1260*. It is described as an educational guide and includes 154 decoupled reactions. As Table 1 shows, 16 of them would hinder the search of any feasible and steady-state pathway. Digging into the constraints of one infeasible LP over *core E.coli*, Equation 17 shows a typical aspect of an infeasible constraint. In this case, infeasible reactions are reactions from an unfolded reversible reaction, as it frequently occurs .

$$succoa_c : -SUCOA_{rev} + SUCOAS + AKGDH = 0 \quad (17)$$

subject to

$$ad1 : SUCOAS > 0$$

$$ad2 : SUCOAS_{rev} = 0$$

$$ad3 : AKGDH \geq 0$$

As it can be inferred, a positive constraint composed only by the reaction *SUCCOAS* drives to an infeasible linear program if its respective reverse reaction *SUCOAS_Rev* is forced to have null flux rate. The way to know if a reaction is well represented in the metabolic reconstruction studied is to force its reverse reaction to be absent. In this case the reaction *SUCCOAS* is not well represented or its presence cannot be used by LP techniques. Only its reverse *SUCCOAS_rev* will be found in extracted EFMs or pathways by any method based on LP.

Fig. 2. Graph showing the infeasible reaction *SUCOAS*

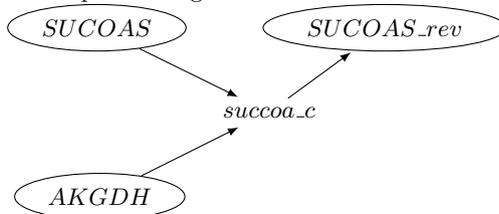


Figure 2 shows the graph representation of the Equation 17 where the unique infeasibility of the reaction *SUCCOAS* appears. Table 2 shows the full set of infeasible reactions from *core E. coli metabolic model*.

Subset of core Ecoli reactions
<i>SUCCOAS</i>
<i>EX_fru(e)</i>
<i>EX_fum(e)</i>
<i>EX_glc(e)</i>
<i>EX_gln_L(e)</i>
<i>EX_mal_L(e)</i>
<i>EX_o2(e)</i>
<i>EX_pi(e)</i>
<i>RPI</i>
<i>ACONTa_rev</i>
<i>ACONTb_rev</i>
<i>ADK1_rev</i>
<i>EX_for(e)_rev</i>
<i>G6PDH2r_rev</i>
<i>O2t_rev</i>
<i>PIt2r_rev</i>

Table 2. Infeasible reactions included in the core Ecoli model

We have analyzed also the stoichiometric matrix used in [17] and 6 of the 33 decoupled reactions are infeasible. The time and resources saved in the worst case can be potentially considerable in any size of stoichiometric matrix. The higher is $|I|$ the higher are the computational/time savings.

Our proposal uses linear programming to detect infeasible reactions, but this set of infeasible reactions can be used by any of the current methods to extract EFMs or pathways even if they do not rely on linear programming as calculation strategy. Obviously, if the method relies on linear programming the benefits can be translated into saved invocations to its respective linear programming solver.

5 Results

Assuming no knowledge at all about I , the probability of building a feasible A with k reactions is the same as the one of choosing k feasible reactions from R .

Figure 3 shows the probability expressed in Equation 15 depending on the size of the additional constraint k . It also shows the impact of using R instead of R' for the *iAF1260* and *core E.coli* reconstructions respectively to illustrate the results. The optimal case would be probability 1 but this is only possible when we chose all the constraints from R' .

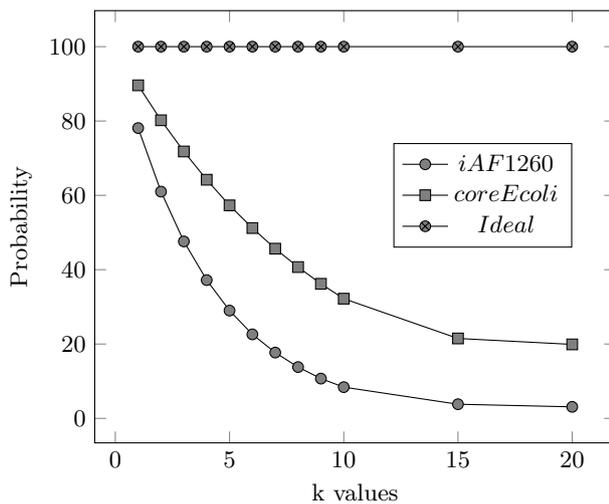


Fig. 3. Probability of getting a feasible LP using a random additional constraint with k reactions. iAF1260 has the worst probability.

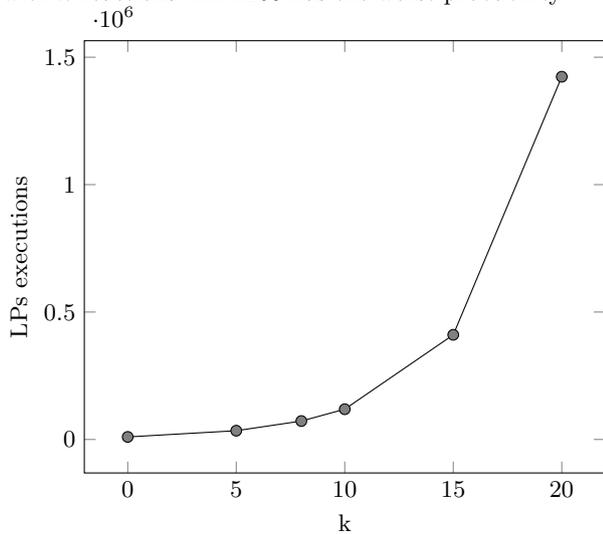


Fig. 4. Amount of solved LPs to get 10000 pathways or EFMs considering the size k of the set of additional constraints for iAF1260 reconstruction.

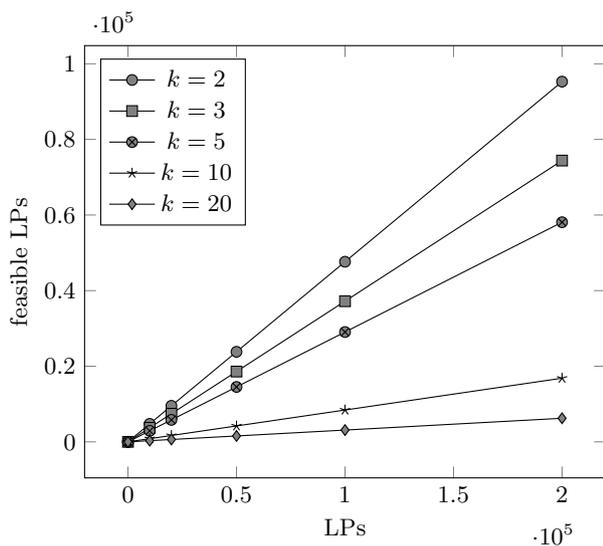


Fig. 5. Obtained feasible LPs (y-axis) after different amount of LPs execution, given a randomly built set of constraints with k reactions for the iAF1260 reconstruction.

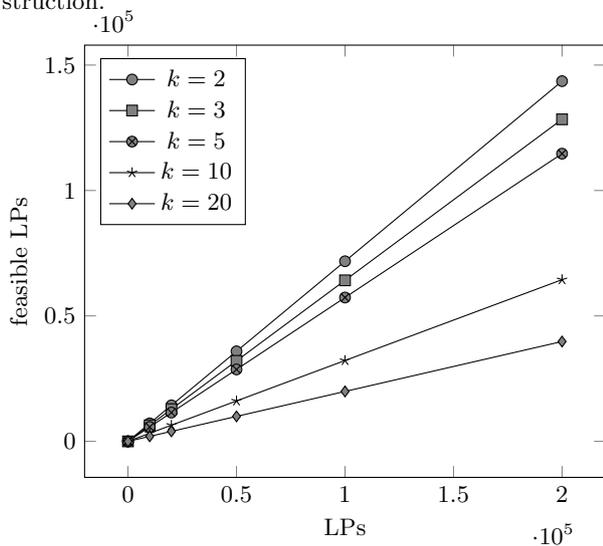


Fig. 6. Obtained feasible LPs (y-axis) after different amount of LPs execution, given a randomly built set of constraints with k reactions for the core E.coli reconstruction.

Figure 4 the amount of solved LPs needed to get 10000 pathways or EFMs considering the size k of the set of additional constraints for iAF1260 reconstruction. If $k \geq 20$, more than 30 times more LPs have to be executed to get 10000 feasible solutions.

Figure 5 shows the linear relationships representing the rate of obtained feasible LPs with the decreasing slopes depending on k for iAF1260 reconstruction. A small k could give acceptable rate of feasible LPs with constraints from $|R|$. As the subset of reactions taken from R raises, the slopes and the probability of building a feasible LP problem decreases.

Figure 6 illustrates the same behavior for the *core E.coli* reconstruction. The comparison between both figures shows that the slope decreases more when the ratio $|R'|/|R|$ gets lower.

Theoretical calculations can be simulated by an asymptotically long enough execution of one hypothetical program to search pathways generating randomly huge amounts of possible sets of additional constraints (and thus LPs) in R . The simulation of experiment over R and R' separately is done using random constraints so the feasible problems are distributed evenly between the infeasible ones.

Depending on the method used to generate the additional constraints the computational savings can vary during the execution, but if theoretically we would pass through all the possible combinations the saved efforts can be substantial using R' . What is clear is that is preferable to explore solutions among R' that only contains feasible reactions to be able to solve mathematically biological questions.

6 Conclusion

Metabolic reconstructions are made to model the cell using the available and proven data and also keeping some reactions or metabolites to provide the most real scenario as possible. Some inclusions are required to adjust thermodynamically the reactions and some other parts may be incomplete due to the fact that the reconstruction of the metabolic map is a simplification of the full map or because the entire map is not fully known. From a biological point of view, any reaction that has been included in a metabolic reconstruction is feasible under the proper balance conditions.

Metabolic infeasibility refers to the impossibility to get an scenario where some metabolic reactions occur against their thermodynamic right sense. Linear programming infeasibility refers to the impossibility of getting a compatible solution with all the added constraints. This is a mathematical issue. To be rigorous, the presence of incompatible built-in constraints joined to the primary constraints must be mathematically checked. Although finding LP infeasibilities does not mean anything against the biological validity of the reconstruction of reference, it can suppose a big difference from the point of view of time and resources that require the pathway extraction methods.

The computational cost associated to the task of retrieving pathways and EFMs en GSMNs can be extremely high when considering large networks. Therefore, methods to alleviate this computational effort can help to extract meaningful biological information from these networks in reasonable time. Here we have proposed a method that accounts for this problem. Avoiding infeasibilities is a key point in every pathway extraction method for GSMNs. Our results permits to do a previous curation of the stoichiometric matrix that assures that no infeasible problems will be formulated when the approach chosen is linear programming. This proposal can be viewed as a previous stage for any existing method and can help to simplify their implementation improve their efficiency.

The benefits of our proposal are the savings of computational resources reducing the amount of LP executions, the floating-point calculations, the amount of variables in the proposed linear program and finally reducing drastically the mean time per LP successful solution. Besides, it cleans out the stoichiometric useless data from the point of view of linear programming and anticipates if one biological question can be answered with the used stoichiometry.

References

1. Masanori Arita. Metabolic reconstruction using shortest paths. *Simulation Practice and Theory*, 8(1):109–125, 2000.
2. Anthony P Burgard, Priti Pharkya, and Costas D Maranas. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnology and bioengineering*, 84(6):647–657, 2003.
3. Jose Francisco Hidalgo Céspedes, Francisco De Asís Guil Asensio, and Jose Manuel García Carrasco. A new approach to obtain efms using graph methods based on the shortest path between end nodes. In *Bioinformatics and Biomedical Engineering*, pages 641–649. Springer, 2015.
4. Adam M Feist, Christopher S Henry, Jennifer L Reed, Markus Krummenacker, Andrew R Joyce, Peter D Karp, Linda J Broadbelt, Vassily Hatzimanikatis, and Bernhard Ø Palsson. A genome-scale metabolic reconstruction for escherichia coli k-12 mg1655 that accounts for 1260 orfs and thermodynamic information. *Molecular systems biology*, 3(1):121, 2007.
5. J Forrest. Clp-coin-or linear program solver. In *DIMACS Workshop on COIN-OR*, July, pages 17–20, 2006.
6. Peggy P Hsu and David M Sabatini. Cancer cell metabolism: Warburg and beyond. *Cell*, 134(5):703–707, 2008.
7. IBM. IBM ILOG CPLEX Optimizer.
url<http://www-01.ibm.com/software/integration/optimization/cplex-optimizer/>, Last 2010.
8. Christoph Kaleta, Luis Filipe de Figueiredo, Jörn Behre, and Stefan Schuster. Efmeflower: Computing elementary flux modes in genome-scale metabolic networks. In *Lecture Notes in Informatics-Proceedings*, volume 157, pages 179–189, 2009.
9. Hans H Kim, Hyun Joo, Tae-Ho Kim, Eui-Yong Kim, Seok-Ju Park, Ji-Kyoung Park, and Han-Jip Kim. The mitochondrial warburg effect: a cancer enigma. *Interdisciplinary Bio Central*, 1(2):7–1, 2009.

10. Steffen Klamt and Jörg Stelling. Combinatorial complexity of pathway analysis in metabolic networks. *Molecular biology reports*, 29(1-2):233–236, 2002.
11. Eugene L Lawler. *Combinatorial optimization: networks and matroids*. Courier Corporation, 1976.
12. Jeffrey D Orth, Ronan MT Fleming, and Bernhard Ø Palsson. Reconstruction and use of microbial metabolic networks: the core escherichia coli metabolic model as an educational guide. *EcoSal Plus*, 4(1), 2010.
13. Jon Pey and Francisco J Planes. Direct calculation of elementary flux modes satisfying several biological constraints in genome-scale metabolic networks. *Bioinformatics*, page btu193, 2014.
14. Jon Pey, Juan A Villar, Luis Tobalina, Alberto Rezola, José Manuel García, John E Beasley, and Francisco J Planes. Treeefm: calculating elementary flux modes using linear optimization in a tree-based algorithm. *Bioinformatics*, page btu733, 2014.
15. Jennifer L Reed, Thuy D Vo, Christophe H Schilling, Bernhard O Palsson, et al. An expanded genome-scale model of escherichia coli k-12 (ijr904 gsm/gpr). *Genome Biol*, 4(9):R54, 2003.
16. Jan Schellenberger, Nathan E Lewis, and Bernhard Ø Palsson. Elimination of thermodynamically infeasible loops in steady-state metabolic models. *Biophysical journal*, 100(3):544–553, 2011.
17. Stefan Schuster, Thomas Dandekar, and David A Fell. Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. *Trends in biotechnology*, 17(2):53–60, 1999.
18. Stefan Schuster and Claus Hilgetag. On elementary flux modes in biochemical reaction systems at steady state. *Journal of Biological Systems*, 2(02):165–182, 1994.
19. Stefan Schuster, Claus Hilgetag, John H Woods, and David A Fell. Reaction routes in biochemical reaction systems: algebraic properties, validated calculation procedure and example from nucleotide metabolism. *Journal of mathematical biology*, 45(2):153–181, 2002.
20. Evgeni Selkov, Natalia Maltsev, Gary J Olsen, Ross Overbeek, and William B Whitman. A reconstruction of the metabolism of methanococcus jannaschii from sequence data. *Gene*, 197(1):GC11–GC26, 1997.
21. Matthew G Vander Heiden, Lewis C Cantley, and Craig B Thompson. Understanding the warburg effect: the metabolic requirements of cell proliferation. *science*, 324(5930):1029–1033, 2009.
22. Otto Warburg, Franz Wind, and Erwin Negelein. Ueber den stoffwechsel von tumoren im körper. *Journal of Molecular Medicine*, 5(19):829–832, 1926.

SalivaPRINT Toolkit: Development, Challenges and Applications

Igor Cruz¹, Eduardo Esteves¹, Mónica Fernandes¹, Nuno Rosa¹, Maria Correia¹,
Marlene Barros¹

¹ Universidade Católica Portuguesa, Center for Interdisciplinary Research in Health (CIIS),
Institute of Health Sciences (ICS), Viseu, Portugal

{Address correspondence to: Marlene Barros, PhD, Director of Center for Interdisciplinary
Research in Health (CIIS), Senior Scientist at SalivaTec Universidade Católica Portuguesa,
Estrada da Circunvalação 3504-505 Viseu – Portugal; Tel. +351232430200 - Fax
+351232428344 email: mbarros@viseu.ucp.pt}

Abstract. Over the last decade saliva has been increasingly regarded as a source of nucleic acids and proteins, which can be used as biomarkers, both from the host and the microbiome. Major advances in capillary electrophoresis performed in automated systems enable an efficient separation of proteins in reduced sample volumes and can be easily applied to saliva samples. This strategy allows the establishment of individual profiles with characteristic patterns reflecting each individual molecular phenotype - SalivaPrint. The premise is that some physiological or pathological conditions may be identified by specific protein profiles, reflecting dysregulated molecular mechanisms. The analysis of the information from SalivaPRINT allows for the identification of salivary profiles characteristic of different physiological conditions.

To date, there were no computational strategies that would systematically and automatically collect, integrate and visualize capillary electrophoresis data. In this work we present SalivaPRINT Toolkit, a computational strategy to analyze SalivaPRINT data, demonstrating some of its applications and identification of the improvements needed. By retrieving information from the data present in multiple profiles and analyzing them with computational strategies it is possible to address different health status dimensions. By comparing individuals affected with a particular condition or pathology with healthy standards, SalivaPRINT Toolkit can identify the altered proteins and provide **molecular clues** to which are the underlying dysregulated pathways or mechanisms involved in each condition. Furthermore, by collecting and analyzing samples from different stages of a particular disease and using clustering techniques, it is possible to achieve a better understanding of the pathologic spectrum, co-morbidities and its evolution over time, allowing **patient stratification** and leading to a **personalized medicine** approach. The ultimate goal of SalivaPRINT Toolkit is to support a non-invasive, easy and cost-effective diagnostic method. This work communicates recent discoveries of the SalivaTec research group (<http://salivatec.viseu.ucp.pt>) using salivary electrophoretic profiles applied to different conditions such as **type 2 Diabetes Mellitus, Periodontitis and Coeliac disease**.

Keywords: Data analysis, Salivary Proteins; Protein phenotypes; Protein profiling

When intrinsically disordered proteins are ordered

Erzsébet Fichó¹, István Simon¹ and Bálint Mészáros²

¹ Protein Structure Research Group, Institute of Enzymology, RCNS, Hungarian Academy of Sciences, Magyar Tudósok krt 2, Budapest, H-1117 Hungary

² MTA-ELTE Lendület Bioinformatics Research Group, Department of Biochemistry, Eötvös Loránd University, Pázmány Péter sétány 1/c, Budapest, H-1117 Hungary

ficho.erszebet@ttk.mta.hu

Abstract. It was commonplace to say that a protein needs a structure to function. But in the last two decades it became clear that a biological function doesn't necessarily require a stable structure. While ordered proteins have a well-defined tertiary structure on their own, intrinsically disordered proteins (IDPs) are highly flexible, fluctuating between many alternative conformations. Their disordered nature enables them to fulfill several vital biological roles in the living cell, they participate in transcription, cell signaling, regulation, and stress-response. Disordered proteins rarely act alone; they are key elements of protein-protein interaction networks, often playing crucial roles in signal transduction. In recent years it became clear that in line with their essential functions, many IDPs are involved in disease development.

Protein complexes formed by ordered proteins have been well studied in the past. The growing number of known disordered proteins and their functions motivate us to analyse interactions in complexes, where either one, or all participating protein partners are disordered. Protein complexes formed exclusively by disordered proteins were a little bit of grey of protein interactions. These interactions are unique because the interacting complexes are ordered, while all participating proteins are unstable when separated (through a process termed mutual synergistic folding (MSF)).

Although these interactions are vital for the living cells and some solved MSF protein complex structures are quite well known, we need an orders of magnitude larger dataset to analyse these type of interactions. Mutual Folding Induced by Binding (MFIB) - available at <http://mfib.enzim.ttk.mta.hu/> - is a repository for protein complexes that are formed exclusively by IUPs. As these proteins lack a well-defined 3D structure in their monomeric form, their folding is induced by the assembly of the complex. MFIB contains over 1400 complex structures organized into 205 entries. These entries came from all three domains of life and also include complexes from viral proteins shedding light on the importance of synergistic folding in host-pathogen interactions.

In order to understand and to analyse interactions formed by IUPs, we also need the Disordered Binding Site (DIBS) database - available at <http://dibs.enzim.ttk.mta.hu/>. DIBS is a repository for protein complexes that are formed between IDPs and globular/ordered partner proteins (coupled-folding-and-binding). DIBS contains more than 1500 protein complexes grouped to 772

entries. This repository also provides the dissociation constants of their interactions, as well as the description of potential post-translational modifications modulating the binding strength and linear motifs involved in the binding.

With the previous studies of globular proteins and using information from MFIB and DIBS, for the first time we are able to get a full view of the entire spectrum of the IDP interactome. In this talk we would like to highlight the differences between various types of interactions mediated by IDPs - in terms of sequence, structure, function and its regulation. This work can be the cornerstone of the basic understanding of how the interplay between protein folding and interaction modulates critical properties of the resulting complexes.

Keywords: disordered protein; mutual synergistic folding; protein-protein interaction; coupled folding and binding

The binding of agonists to the NOP receptor revealed “in silico” by Multi-flexible docking and QM/MM.

Stefano della Longa^{1,} and Alessandro Arcovito²*

¹Department of Life, Health and Environmental Sciences, University of L'Aquila, L'Aquila, Italy; and

²Institute of Biochemistry and Clinical Biochemistry, Catholic University of Sacred Heart, Rome, Italy

KEYWORDS: AT-127, Molecular dynamics, GPCR, opioid receptor, nociception receptor, nopr agonists, oprx, N/OFQ

Nociceptin/orphanin FQ (N/OFQ) is a peptide involved in the activation of a cellular pathway resulting in the inhibition of cAMP synthesis and in the increase of membrane permeability to K⁺. The nociceptin receptor (NOP) activation controls biological functions and physiological effects noted in the nervous system (central and peripheral), the cardiovascular system, the airways, the gastrointestinal tract, the urogenital tract and the immune system spectrum (1-3).

Non peptide antagonists of the nociceptin receptor (NOP) are widely studied for their possible clinical use as antidepressant drugs (4), and recently, the structures of NOP in complex with some piperidine-based antagonists have been revealed by X-ray crystallography, giving outstanding informations on the role of some key residues for the affinity and the selectivity profile of those molecules. On the other hand, selective non peptide NOP agonists have shown significant efficacy for anxiety and pain(5), but the structure of activated NOP is still not known apart from “in silico” studies(6).

In this study, a Multi-Flexible docking procedure, i.e. docking to multiple receptor conformations, together with hybrid quantum mechanics / molecular mechanics (QM/MM) simulations have been carried out to provide the binding mode of some novel NOP agonists, to be compared with the binding mode of antagonists. According to our results, and in analogy with what is known from the comparison of the crystal structures of the inactive and active μ -opioid receptor, the overall structural differences in the orthosteric binding pockets are rather small. In the case of piperidine-based agonist compounds (e.g. compound AT-127, Fig. 1 and Fig. 2, orange sticks), the pivotal role of residue D130^{3,32} (upper indexes are Ballesteros-Weinstein notations) is analogous to that of the already known X-ray structures of antagonist-bound opioid receptors. The comparison with the crystal structure of inactive NOP in

complex with to compound C24 (Fig. 2, black wires) reveals that the agonist AT-127 can be accommodated in a receptor conformation having a small outward movement of helix 6. In this conformation the aromatic side chain of residue W276^{6,48} has moved toward the core triad formed by F272^{6,44}, P227^{5,50} and T138^{3,40}. This early movement, could be associated to the destabilization of the core conformation, that in turn, in the presence of the agonist in the orthosteric pocket and the G-protein at the cytoplasmic side, triggers the receptor toward an activated conformation.

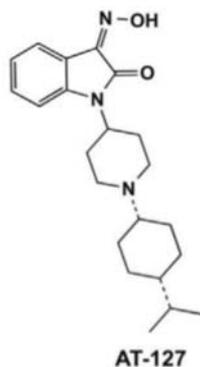


Fig. 1

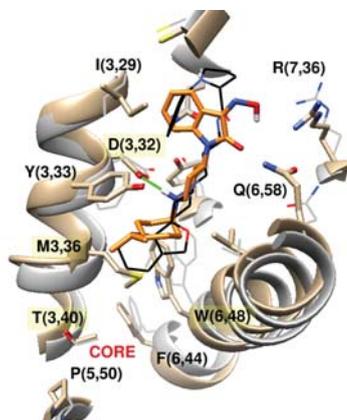


Fig. 2

REFERENCES

1. Lambert, D. G. (2008) The nociceptin/orphanin receptor: a target with broad therapeutic potential. *Nature Rev. Drug Discovery* **7**, 694-710
2. Chiou, L. C., Liao, Y. Y., Fan, P. C., Kuo, P. H., Wang, C. H., Riemer, C., and Prinssen, E. P. (2007) Nociceptin/orphanin FQ peptide receptors: pharmacology and clinical implications. *Curr Drug Targets* **8**, 117-135
3. Mogil, J. S., and Pasternak, G. W. (2001) The molecular and behavioral pharmacology of the orphanin FQ/nociceptin peptide and receptor family. *Pharmacol Rev* **53**, 381-415
4. Gavioli, E. C., and Calo, G. (2013) Nociceptin/orphanin FQ receptor antagonists as innovative antidepressant drugs. *Pharmacol Ther* **140**, 10-25
5. Toll, L., Bruchas, M. R., Calo, G., Cox, B. M., and Zaveri, N. T. (2016) Nociceptin/Orphanin FQ Receptor Structure, Signaling, Ligands, Functions, and Interactions with Opioid Systems. *Pharmacol Rev* **68**, 419-457
6. Della Longa, S., and Arcovito, A. (2016) A Dynamic Picture of the Early Events in Nociceptin Binding to the NOP Receptor by Metadynamics. *Biophys J* **111**, 1203-1213

Classification of cardiovascular pathologies in artificial signals of a lumped parameter model using a naive Bayes algorithm

Stefan Krickl^{1,2}, Manuel Debic¹ and Stefan Bernhard^{1,2}

¹ Pforzheim University of Applied Sciences,
Tiefenbronnerstr. 65, 75175 Pforzheim, Germany

² Fachbereich Mathematik, Freie Universität Berlin,
Germany

{stefan.krickl,manuel.debic,stefan.bernhard}@hs-pforzheim.de
<http://www.hs-pforzheim.de>

Abstract

Pathological bulges of the blood vessels, especially aneurysms in the abdominal (abdominal aortic aneurysm, AAA) or thoracic aorta (thoracic aortic aneurysm, TAA) are a highly underestimated problem. According to Criqui et. al. these diseases affect 12 to 14 % of the population. However the asymptomatic course of the diseases in most patients and the misinterpretation of the initial symptoms, in about one third of the symptomatic patients, lead to a categorical underdiagnosed clinically situation. A supplementary non-invasive method, which is independent of the examiners experience, would be highly desirable for early diagnosis at the family physician level. The aim of this study is the development and validation of a methodology to classify the above mentioned vascular pathologies using non-invasive cardiovascular pressure and flow measurements.

For this purpose, we studied the classification performance of a naive Bayes classifier, based on pressure and flow signal intervals and amplitudes. The features were extracted from artificial pressure-flow signals generated by the numerical cardiovascular modeling tool SISCA, that describes the blood flow in the cardiovascular network based on the theory of Westerhofs et. al. in a zero dimensional lumped parameter approach. The model under consideration contains 123 segments, including 25 terminal segments. The boundary conditions at the inlet segment is driven by an input pressure time series within the pressure range of 80 to 120 mmHg and a heartrate of 75 bpm. The blood viscosity was chosen to be 4 mPa s, the lumped elements are modeled using a linear visco-elastic behaviour. For each segment, there are two time series, one for the pressure and one for the flow, leading to a total number of $2 \cdot N_s = 2 \cdot 123 = 246$ artificial pressure flow signals, where N_s is the total number of segments. The network equations are solved using the lsim Matlab (Mathworks) solver within the SISCA simulation tool published by Huttary et. al.

Besides the control group of normal patients, the variational scenario is based on two pathological conditions of the AAA and TAA with different extension and location. The nominal cross-sectional areas were enlarged between 200 % and 600 %, while the length of the aneurysms were modified within a range of 30 and 90 mm. Segments of the aorta with a cross sectional area of less than 300 % are considered small, between 300 and 400 % are considered medium and greater than 400 % are considered as large aortic aneurysm. The model parameters within the pathology Segments were varied randomly within their uncertainty ranges using the Monte-Carlo method. Finally the pressure and flow waveforms were collected at carefully chosen locations in the left leg/ arm using a global sensitive analyses method presented by Gul et. al.

The high parameter dimensionality and hence the costly experimental situation was optimised using the design of experiments (DOE) method, so all variables were considered in a minimal number of experiments, and the result was verified by the bootstrap method to find a 95 % confidence interval of each parameter for each state variable, that lies within ± 5 % uncertainty as compared to the mean value of all simulations. To cover the entire range of parameter spaces with reasonable computational cost, Latin hypercube sampling (LHS) was used. The minimum number of simulations for each parameter to achieve uncertainty in the main sensitivity index below ± 2 % uncertainty as compared to the mean value of all simulations, is satisfied for values greater 3500.

The classification features were preselected according to the optimal measuring sites for the disease obtained by sensitivity analysis proposed by Gul et. al and by a combination of the variance and covariance of the measurement set. Depending on the number and type of pathologies, a combination of several measuring points is useful to improve the classification quality. Detection and classification with this preselection method feature matching performs extremely rapid by using the variance and variance covariance matrix. Additionally the feature extraction and dimensionality reduction was effectively performed by using Principal Component Analysis (PCA).

The classification performance was evaluated using sensitivity, specificity, classification accuracy and mean squared error (MSE). Depending on disease severity, the procedure achieved a global testing classification accuracy of over 90%.

We assume that the method will improve of the cardiovascular signals in the early diagnosis of diseases in near future in a diagnostic decision support system. The naive Bayes classifier has shown good results for artificial signals, however it is planned to apply the method to clinical data in near future.

Keywords: Classification, naive Bayes, cardiovascular pathologies, aneurysms, pressure-flow signal, design of experiments, feature selection, bootstrap, Monte Carlo, Latin hypercube sampling, sensitivity analysis

Putative oxidative folding pathways of the disulfide rich inhibitory peptide Tridegin

*Ajay Abisheck Paul George, Pharmaceutical Biochemistry and Bioanalytics,
Pharmaceutical Institute, University of Bonn, Bonn, Germany*
*Arijit Biswas, Institute of Experimental Hematology and Transfusion Medicine,
University hospital Bonn, Bonn, Germany*
*Monica Sudarsanam, Pharmaceutical Biochemistry and Bioanalytics, Pharmaceutical
Institute, University of Bonn, Bonn, Germany*
*Diana Imhof, Pharmaceutical Biochemistry and Bioanalytics, Pharmaceutical Institute,
University of Bonn, Bonn, Germany*

Keywords

Tridegin – oxidative folding - molecular dynamics – factor XIIIa

The oxidative folding pathway of disulfide peptides has been described by two extreme modes or pathways with a large group of proteins falling in between these extremes[1,2]. In such a situation, it makes sense to exhaustively sample the conformations visited by such cysteine rich peptides to make meaningful deductions. We chose Tridegin, a 66-mer peptide as candidate given its biological relevance in blood coagulation[3,4] and its ability to form three different disulfide bonded isomers[5] with one of its disulfide combination (between C19 and C25) being consistent amongst all three isomers. We have started an in silico oxidative folding experiment of this peptide using successive differently scaled all atom molecular dynamics simulation. Since no resolved structures are present for any of the three isomers, threaded optimized models of the same serve as primary end points to the folding process. Currently, a run of ~3 μ s shows the appearance of short helices observed in all three isomer models. Interestingly the C19 and C25 residues appear to approach each other (presently 6 Å distance between S γ atoms) in the manner of generating native disulfide bonds after overcoming side chain interaction thresholds. On a significantly longer time-scale of simulation we would eventually like to find out if the natively folded consistent disulfide bond acts as master regulator for all downstream folding in this protein.

References

1. Chang JY. Diverse pathways of oxidative folding of disulfide proteins: Underlying causes and folding models. *Biochemistry*. 2011;50: 3414–3431. doi:10.1021/bi200131j
2. Chang J-Y. Distinct Folding Pathways of Two Homologous Disulfide Proteins: Bovine Pancreatic Trypsin Inhibitor and Tick Anticoagulant Peptide. *Antioxid Redox Signal*. 2011;14: 127–135. doi:10.1089/ars.2010.3634
3. Seale L, Finney S, Sawyer RT, Wallis RB. Tridegin, a novel peptidic inhibitor of factor XIIIa from the leech, *Haementeria ghilianii*, enhances fibrinolysis in vitro. *Thromb Haemost*. 1997;77: 959–963.
4. Finney S, Seale L, Sawyer RT, Wallis RB. Tridegin, a new peptidic inhibitor of factor XIIIa, from the blood-sucking leech *Haementeria ghilianii*. *Biochem J*.

- 1997;324 (Pt 3: 797–805.
5. Böhm M, Bäuml CA, Hades K, Steinmetzer T, Roeser D, Schaub Y, et al. Novel insights into structure and function of factor XIIIa-inhibitor tridegin. *J Med Chem.* 2014;57: 10355–10365. doi:10.1021/jm501058g

Red blood cell model with different implementation of viscoelastic parameter

Mariana Ondrušová*

University of Žilina, Faculty of Management Science and Informatics, Department of
Software Technology,
mariana.ondrusova@fri.uniza.sk,
WWW home page: <http://cell-in-fluid.fri.uniza.sk>

Abstract. Dynamic properties of red blood cells are closely related to their elasticity parameters. Elasticity parameters influence stretching and the return to original shape. With a correctly chosen simulation model we are able to approximate the results of biological experiments. Elasticity parameters are used to preserve a triangular grid, which is the base for our spring-network model. The elastic properties of our model are provided by six elastic moduli.

One of the parameters is viscosity. Viscosity influences dynamic of the experiments during cell movement and it also influences the change of the cell shape. Viscosity can be implemented in the computational model in two ways. Both methods work with the size of the edge of the triangular grid between two points and the difference of velocities of these two points. In the first method, viscous force is calculated using the difference of the velocities of the two points. The second method projects the resulting difference a perpendicular on the edge of the two points. Resulting force is then calculated from the projected velocity. Results were obtained with a suitably chosen simulation model. These results indicate a difference between the specific implementations of the viscous parameter. The following chapters reflect work in progress.

Keywords: red blood cell, viscous parameter, computational modeling, simulation

1 Introduction

Microfluidic simulations are based on interactions between objects representing red blood cells in the fluid flow [1, 2].

Flow of liquid: Based on the lattice-Boltzman method, it is assumed that the space that fills the liquid is divided into evenly distributed points in the cubic network. This grid is fixed throughout the simulation time. More details can be found in [3].

* This work was supported by the Ministry of Education, Science, Research and Sport of the Slovak Republic under the contract No. VEGA 1/0643/17 and by the Slovak Research and Development Agency under the contract No. APVV-15-0751

Object representing the red blood cell: The object is described by points that together form a triangular network representing the surface of the red blood cell. Red blood cell elasticity assured by six elastic moduli - stretching, bending, local and global area, volume and viscous modulus. The main idea for them was taken from [4]. The first 5 moduli were adapted for our needs in [5]. The viscous modulus was taken from [6]. The first 5 elastic moduli maintain the cell membrane shape. The viscous modulus complements other moduli, but its primary role is not shape preservation. The force generated by viscoelastic modulus is opposite to the force acting during the change of the cells form. Specifically, the faster the cell shape changes, the greater is the viscous force acting in the opposite of this change. The force is calculated as

$$F_{visc}(A, B) = k_{visc} \frac{\Delta L_{ab}(t)}{\Delta t} n_{AB} \quad (1)$$

where k_{visc} is viscoelastic coefficient and $\frac{\Delta L_{ab}(t)}{\Delta t}$ is speed of deformation and n_{AB} is vector determining the direction of the force. More details can be found in [7, 8].

Fluid-Object Coupling: Describes movement of objects in fluid. Immersed objects are characterized by boundary points. The movement of these points originates from the interaction between the fluid force and the boundary points and vice versa. Speed of a boundary point is the same as the fluid velocity that is recalculated at the same point.

2 Viscosity

When a cell moves in dynamic systems, the cell membrane is torn and deformed. Dynamic systems include all options when the cell moves, rolls or changes shape. Deformation of cells is characterized by elastic and viscous properties. They define the direction in which the membrane bends, stretches or shrinks. Viscosity always acts against this change and thus slows down the deformation of the membrane. It is important to remember, that viscosity is only relevant in the dynamic system, therefore in static experiments such as cell sedimentation, its should be set to zero.

Viscosity can be implemented into the model by two principles. Both of them we can found in [9].

1. principle: Figure 1 shows two points of the network connected by a common edge L, moving at velocities v_1 and v_2 . In general, these velocities are not equal. The relative velocity of the first point with respect to the second point, is given by the velocity difference $v_2 - v_1$. Viscosity acts against this relative velocity. Therefore, viscous force can be defined as a negative multiple of this difference by

$$F_{visc} = k_{visc}(v_2 - v_1). \quad (2)$$

This approach has been used for example in [10].

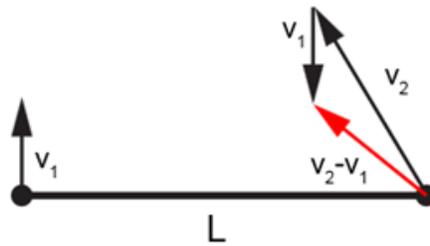


Figure 1. The first principle of viscosity

2. principle: If we consider viscosity as a mechanism which acts only against the change of the edge length, we have to project the relative velocity in the direction of the edge (Figure 2). Therefore, we include projection of the speed in our computations. Meaning that we will velocity project on the edge L . This can be written in the following equation:

$$F_{visc} = k_{visc}(\text{project}_L(v_2 - v_1)). \quad (3)$$

Second principle seems more natural to us and should better reflect the properties of the biological membrane.

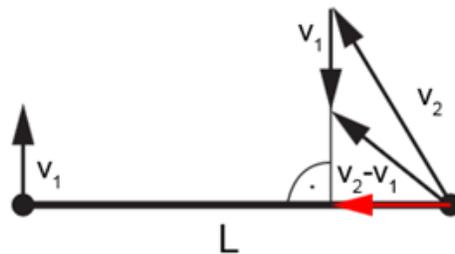


Figure 2. The second principle of viscosity

3 Current simulations

We verified the influence of the membrane viscosity on two simulations. Both of them were performed in a cubic box, and we used a sphere-shaped object situated in a centre of the channel. In the first simulation, we chose two points

with a common edge that show the highest rotation period during the simulation. Thus, the points whose rotation frequency is as large as possible. In the second simulation, we have chosen points that have the lowest possible rotation frequency. During rotation, these points move at a given velocity and direction. For the first pair, the vectors of velocities are quite comparable (Figure 1,2). For the second pair, the vectors should be opposite. Then the resulting viscous force should be significantly different for the second pair (Figure 3).

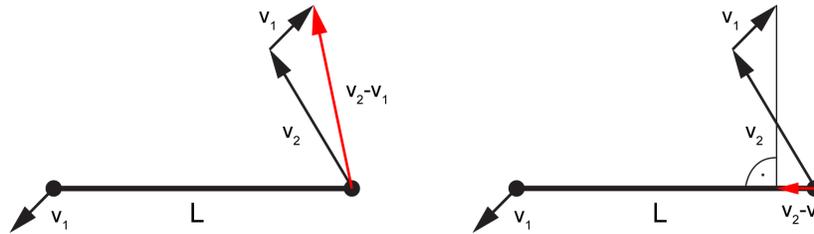


Figure 3. The difference of viscous force in the first and second implementations

As we mentioned before, the viscous force acts during the change of the shape of the cell. The form of the sphere changes only a little during its rotating, so a correctly adjusted viscous force should be small, as well.

References

1. Cimrak, I., Gusenbauer, M., Schrefl, T.: Modelling and simulation of processes in microfluidic devices for biomedical applications. *Computers & Mathematics with Applications* 64.3 (2012), 278-288.
2. Bachrata, K., Bachraty, H.: On modeling blood flow in microfluidic devices. *ELEKTRO 2014: 10th International Conference, IEEE*, (2014), Slovakia, pp. 518-521
3. Dunweg, B., Ladd, A. J. C.: Lattice-Boltzmann simulations of soft matter systems. *Advances in Polymer Science* 221, 89166, (2009)
4. Dupin, M. M.; Halliday, I.; Care, C. M.; et al.: Modeling the flow of dense, Modeling the flow of dense suspensions of deformable particles in three dimensions *Physical review E*, zv. 75, 1. vyd.6, (2007)
5. Jancigova, I., Cimrak, I.: Non-uniform force allocation for area preservation in spring network models. *International Journal for Numerical Methods in Biomedical Engineering*.
6. Busik, M.: Development and optimization model for flow cells in the fluid. [Dissertation thesis] - University of ilina. Faculty of Management Science and Informatics. Department of Software Technology. - Supervisor: doc. Mgr. Ivan Cimrak, Dr. - ilina, FRI ZU, (2017), 114 p.

7. Ondrušová, M., Cimrák, I.: Dynamical properties of red blood cell model in shear flow. (2017). [Online]. Available: <http://cell-in-fluid.fri.uniza.sk/en/content/publications>.
8. Ondrušová, M.: Sensitivity of red blood cell dynamics in a shear flow. Central European Researchers Journal, Vol.3 Issue 1 (2017)
9. Ondrušová, M.: Vývoj, kalibrácia a aplikácia biologických buniek, Žilinská univerzita (2017)
10. Fedosov, D. A., PhD thesis: Multiscale Modeling of Blood Flow and Soft Matter, Brown University, (2010)

Development of multi-agent technology for prediction of the "structure-property" dependence of drugs on the basis of modified algorithms of artificial immune systems

Samigulina Galina and Samigulina Zarina*

Institute of Information and Computing Technologies, Laboratory "Intellectual control systems and forecasting", str. Pushkin 125, 050010, Almaty, Kazakhstan
Kazakh British Technical University, Faculty of Information Technologies, str. Tole be 59, 050000, Almaty Kazakhstan

Due to the rapid development of computer and information technology there has been collected a huge amount of information on the structures of chemical compounds. By 2018, the world's largest repository of chemical information PubChem contains 94,703,175 components, Mol-Instincts - more than 2.85 million components, DETHERM - about 11.3 million of data set, etc. The prediction of the "structure-property/activity" dependence (QSAR) of new drugs with prescribed properties is a multi-stage and expensive procedure, requiring on average about 100 million dollars on the research of one candidate.

The urgent task is the development of new innovative technologies for the computer molecular design of drug compounds based on bioinspired approaches of artificial intelligence. Nowadays, in the world practice there are no universal algorithms for solving this problem, because the results of modeling largely depend on the initial data set. The development of new modified algorithms of Artificial Immune Systems (AIS) [1], based on the immunological reaction of the organism during the invasion of foreign antigens, is also actual. Under the AIS there is meant information technologies that use the concepts of theoretical immunology for solving various applied problems. AIS have the following advantages: memory, training ability, distribution and self-organization, a high degree of parallelism [2].

The developed multi-agent technology for prediction the "structure-property/activity" dependence of drugs allows to choose that algorithm of preliminary data processing (Random Forest, Genetic Algorithm, Gray Wolf Optimization, etc.), which after solving the image recognition problem on the basis of various AIS approaches (Immune Networks Modeling [3,4], Clonal Selection, Negative Selection,) and estimating the generalization error will give the best prognostic result (Fig.1).

Therefore, the proposed multi-agent technology has a set of advantages. This is primarily the ability to solve the problem of image recognition of chemical compounds, that are on the boundary of non-linearly separated classes, and to recognize chemical compounds with an almost identical structure, but having completely different properties.

* Corresponding author. E-mail: zarinasamigulina@mail.ru

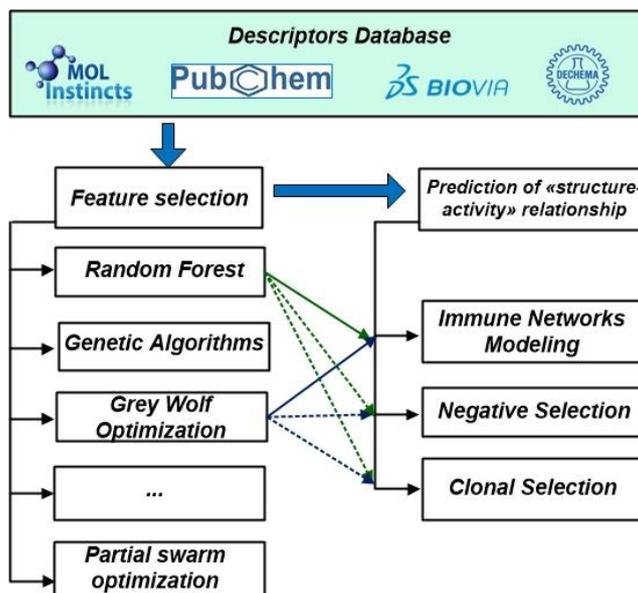


Fig. 1. The technology for prediction of the "structure-property" dependence of drugs based on approaches of artificial immune systems

The work was carried out according to the grant of the KN MES RK on the theme: "Development and analysis of databases for the information system for prediction the "structure-property" dependence of drug compounds based on artificial intelligence algorithms" (2018-2020).

References

1. Read M., Andrews P., Timmis J. An Introduction to Artificial Immune Systems. Handbook of Natural Computing. 1575-1597 (2012).
2. Dasgupta, D., Yu, S., Nino, F. Recent Advances in Artificial Immune Systems: Models and Applications. Applied Soft Computing. 2, 1574-1587 (2011).
3. Samigulina G.A., Samigulina Z.I. Immune network Technology on the Basis of Random Forest Algorithm for Computer-Aided Drug Design //Bioinformatics and Biomedical Engineering. 10208, 50-61 (2017).
4. Samigulina G.A., Samigulina Z.I. Development of the immune network modeling technology of computer molecular design of drugs. Certificate of state registration of intellectual property object in the Committee on Intellectual Property MJ RK. 473 (2011).

Impedance Analysis of Different Shapes of the Normal and Malignant White Blood Cells

Sameh Sherif^{1,2}, Hamdy Abdelhamid², Yehya H. Ghallab^{1,2}, Ahmed Seddik¹, Yehea Ismail².

¹ Biomedical Engineering Department, Helwan University, Cairo, Egypt

² Center of Nano electronics and Devices (CND), Zewail City of Science and Technology and The American University in Cairo (AUC), Giza, Egypt

Email: Sameh.Sherif@aucegypt.edu, {yghallab, habdelhamid,yismail}@zewailcity.edu.eg

Abstract --- In this work, a 3d numerical study for double shell the leukemia cell impedance is presented. A capacitive sensing based technique is used to extract the cell Clausius-Mossotti. The spherical and non-spherical shapes are considered in this work. We found the impedance of non-spherical leukemia shape is higher than the spherical leukemia shape by 10%. A full study of impedance for different shapes of the normal and malignant white blood cell using numerical simulators also is presented. The cell position over the sensing electrodes and their effects also is considered to capture the exact impedance value.

Keywords: Impedance, Leukemia, double-shell, 3d numerical simulation

I. INTRODUCTION

The biological cells carrying different electrical properties can be used to differentiate not only between cell sources but also between the infected and healthy cells that obtained from the same source [1]. Among these electrical properties is the cell impedance [2]. Such instrumentation used to measure the cell impedance is the spectroscopy [3]. Electrical impedance spectroscopy (EIS) is considered a noninvasive method used to analyze single cells, according to their dielectric properties as a function of frequency [4]. Also, EIS measures the cell size or volume at lower frequencies, the cell membrane capacitance, and intracellular features at higher frequencies [5]. The dielectrophoresis [6] and electrophoresis [3] forces are two techniques that used to move the biological cells (in 2D direction or even in the 3D plane) to extract different physical properties for the biological cell such as mass, weight, speed, moving direction and electrical properties. The dielectrophoresis and electrophoresis forces ideas are based on applying non-uniform or uniform electric fields on the cell surface to set up an electronics dipole. The established dipole values are reflecting the cell properties such as their surface area, shape, and internal structure. An interfacing system can be designed to extract the electrical parameters we called it a static technique, i.e. without moving the biological cell such as a capacitive sensing based technique [20]. In [10], they extracted the single shell impedance but ignored the non-spherical shape effects. In our analysis, the double shell model is used to understand the exact physiological phenomena of the white blood cell. [9-11, 15]. Due to the difficulty of introducing an accurate model for the cell (i.e., the shape of the cell influences the accuracy of dielectric characterization). So it's important to introduce an accurate model of the cell at which most of the parameters are included definitely in both the normal and malignant white blood cell to calculate Clausius-Mossotti factor (CM factor) for example as described in [15]. In [4, 9] however, they studied the white blood

cell using the double shell approximation to extract the electrical properties, but they did not consider the cell dimension in their calculations. In [15], they took into account the shape dependence (spherical approximation) in their studies but a single shell approximation is considered and ignoring the effects of the double shell and non-spherical effects.

In this work, we analyzed the double shell approximation of the white blood cell of non-spherical geometric shape.

The paper is organized as; in Section (2) the theory of operation which describes impedance as a function of the voltage applied on sensing electrodes and the current passing through the cell. Section (3) introduces the mathematical expression and different structure of white blood cell and mathematical solution of Clausius-Mossotti factor. Section (4) presents the system structure used to extract the impedance calculations for the normal and malignant white blood cell and the results. The conclusion and remark are presented in Section 5.

II. THEORY OF OPERATION

Figure 1 shows the graphic design and the principles of the microfluidic system .The system is composed of chambers covered two main sensing electrodes which using to be a capacitor with dielectric constant. The cell pushed through the inlet of the chamber to passing through the sensing electrodes and vary the dielectric constant that means the capacitance of electrodes will different that lead to vary the impedance. Using Ohm's law to calculate the impedance (Z) [14]. $Z=V/I$ (1)

Impedance contains both a phase and a magnitude, and it consists of two main components Reactance and conductance which the impedance can represent by the following expression.

$$Z= \sqrt{X_c^2 + R^2} \quad (2)$$

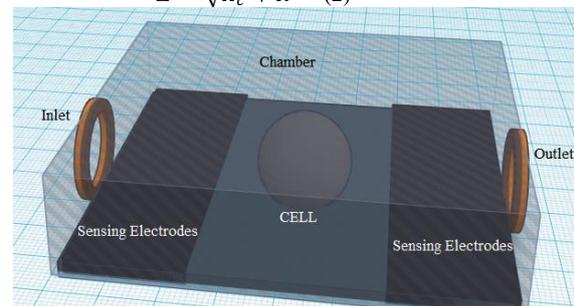


Fig. 1. A 3D graphic Design of microfluidic system

EXTENDED ABSTRACTS

III- MATHEMATICAL ANALYSIS FOR SINGLE AND DOUBLE SHELLS

A- Mathematical Model

The white blood cell as we mentioned in the introduction section can be approximated as a single shell (Fig. 2a) or as a double shell (Fig.2b). In the single shell approximation consists of two homogenous dielectric regions named; the cytoplasm and the cell membrane [8].

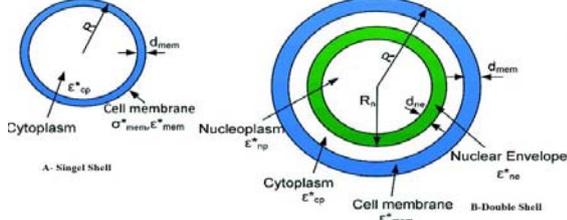


Fig. 2.A Single and double shell models

The Clausius-Mossotti factor (CM factor) of single shell can be represented by [8]:

$$K(w)_s = \frac{w^2(\tau_m \tau_{cm} - \tau_{cp} \tau_{cm}) - 1 + jw(\tau_{cm} - \tau_m - \tau_{cm})}{2 - w^2(\tau_{cp} \tau_{cm} + 2\tau_m \tau_{cm}) + jw(\tau_{cm} + 2\tau_m + 2\tau_{cm})} \quad (3)$$

$$\tau_m = \frac{\epsilon_m R}{\sigma_m d} \quad (4)$$

$$\tau_i = \frac{\epsilon_i}{\sigma_i} \quad (5)$$

Where the Clausius-Mossotti factor (CM factor) for double shell approximation is given by [9],

$$K(w)_D = \frac{(\alpha_{cm} \beta_{cp} + 2\alpha_{cm} \alpha_{ncp} \alpha_{cp} \rho_{12} + \alpha_{cp} \beta_{cm} \beta_{cp} \rho_{23} + \alpha_{ncp} \beta_{cm} \beta_{cp} \rho_{13})}{(\beta_{mcm} \beta_{cp} \beta_{cncp} + 2\beta_{mcm} \alpha_{ncp} \alpha_{cp} \rho_{12} + 2\alpha_{mcm} \alpha_{cncp} \beta_{cp} \rho_{23} + 2\alpha_{mcm} \alpha_{cp} \beta_{cncp} \rho_{13})} \quad (6)$$

$$\alpha_{ij} = \epsilon_i^* - \epsilon_j^* \quad (7)$$

$$\beta_{ij} = 2\epsilon_i^* + \epsilon_j^* \quad (8)$$

$$\rho_{nf} = \left(\frac{R_n}{R_f}\right)^3 \quad (9)$$

Where (m) for medium, (cm) for cell membrane, (cp) for cytoplasm, (nm) for nuclear membrane and (np) for nucleoplasm

TABLE I. DIELECTRIC PROPERTIES OF LEUKEMIA K562, T-NORMAL AND B-NORMAL [15-18]

	Permittivity			Conductivity(S/m)		
	LEUKEMIA	B-Cell	T-Cell	LEUKEMIA	B-Cell	T-Cell
Membrane	8.72	11.1	12.8	$< 10^{-6}$	$27.4 * 10^{-6}$	$56 * 10^{-6}$
Cytoplasm	70.04	60	60	0.503	0.65	1.31
Nuclear envelope	17.05	85.6	17.05	$9 * 10^{-4}$	$9 * 10^{-4}$	$9 * 10^{-4}$
Nucleoplasm	50.12	120	50.12	0.902	1.26	0.902

IWBBIO 2018

B- Clausius-Mossotti factor (CM factor) of Leukemia

The CM factor real and imaginary components of the Leukemia can be calculated from eqs. (3,6). Table 1 lists the main leukemia parameters obtain from [15-18]. Fig. 3 and Fig. 4 represent the mathematical analysis of real and imaginary components of Clausius-Mossotti factor (CM factor) for the double shell of the normal and malignant white blood cell. Fig.5 and Fig.6 show the obtained results Clausius-Mossotti factor (CM factor) for the single shell of the normal and malignant white blood cell.

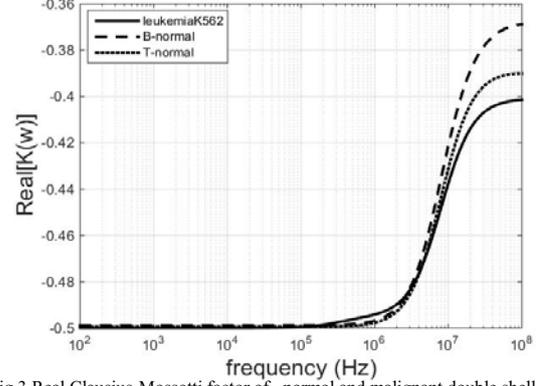


Fig.3 Real Clausius-Mossotti factor of normal and malignant double shell White blood cell.

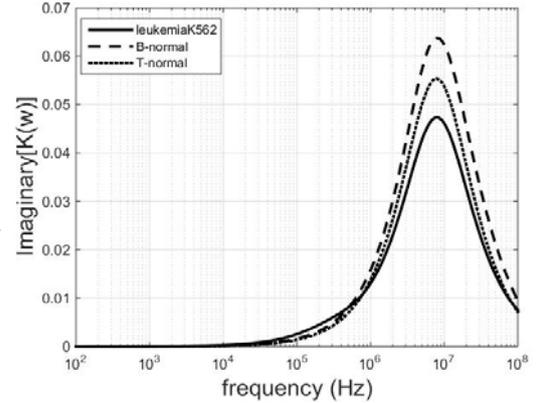


Fig.4 Imaginary Clausius-Mossotti factor of normal and malignant double shell White blood cell.

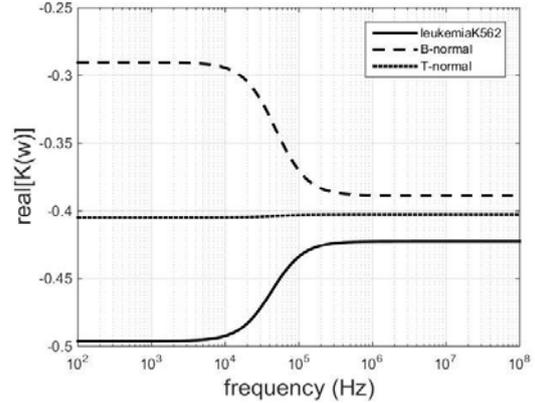


Fig.5 Real Clausius-Mossotti factor of normal and malignant single shell White blood cell.

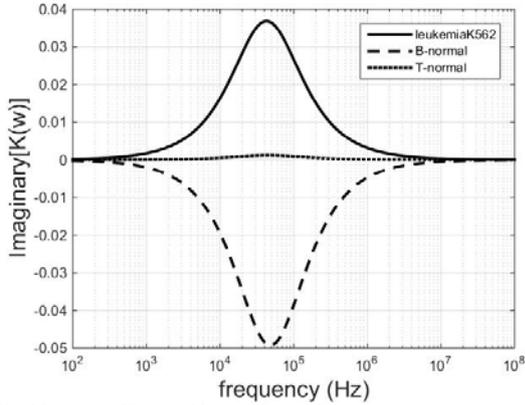


Fig.6 Imaginary Clausius-Mossotti factor of normal and malignant single shell White blood cell

Significant differences can be noticed from Figs. (3-6) between the single and double shell approximations.

IV. 3D NUMERICAL SIMULATION OF SPHERICAL AND NON-SPHERICAL WHITE BLOOD CELL APPROXIMATION.

In this section, the effect of the cell geometry on the impedance will be presented. A two proposed geometries are proposed in this section, spherical and non-spherical geometries as shown in Fig. 7. The simulation results obtained in this section are based on the given parameters in Table 1

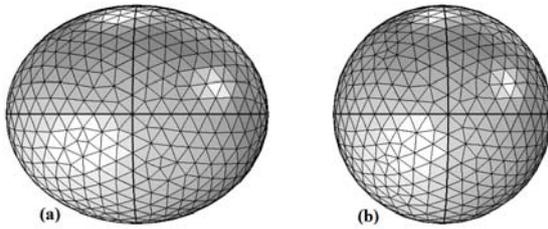


Fig.7 The ellipsoid and Spherical structures for the cell.

A sinusoidal signal of frequencies ranged from 1 KHz up to 10-MHz and peak-to-peak voltage is 1-V applied to the proposed system, Fig. 1. The electric field distribution is shown in Figs. (8-9) all data are analyzed by the COMSOL numerical simulator from 1 KHz to 10 MHz all these data were analyzed by the COMSOL software.

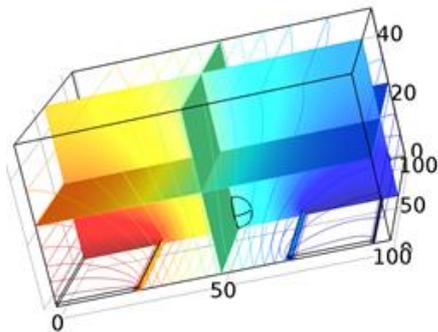


Fig.8 The electric field distribution all dimensions in μm.

Electric Field Vs Distance between Electrodes

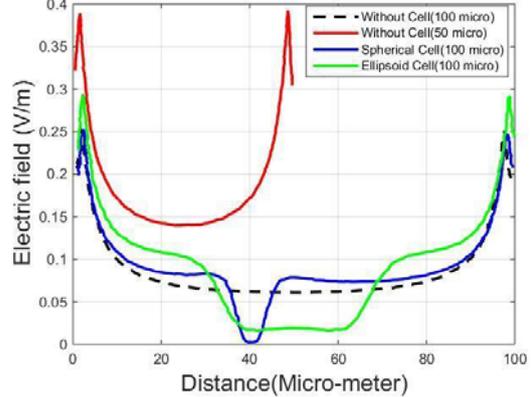


Fig.9 The electric field distribution between electrodes .

A- Impedance analysis

In this section, the variation between the impedance values of white blood cell based on the geometry is presented.

A-1- Spherical Shape

The proposed impedance system, Fig.1 is used to calculate the spherical geometry shape, Fig. 7b. Three different white blood cells are used in this trial with spherical geometries. From Fig. 10, a non-significant difference can be noticed between the three white blood cells (Leukemia, B-Type and T-Type). From Fig. 10, it was also noticed that leukemia has higher impedance than the other two-white blood cells impedance.

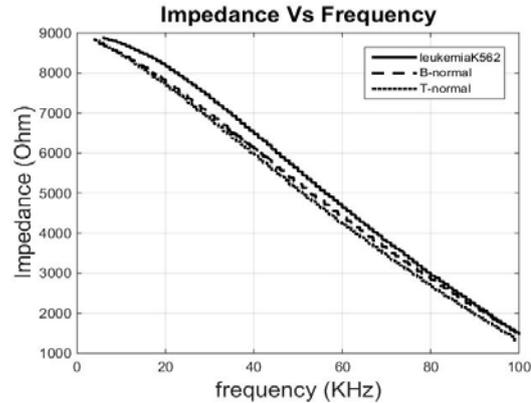


Fig. 10 The impedance of the normal and malignant cell of white blood cell as spherical

A-2- NonSpherical Shape (Ellipsoid shape)

Now applying the ellipsoid shape, Fig. 7b to the proposed system given in Fig. 1. A significant difference between the three given white blood cells is achieved, as shown if Fig. 11. This gives us a very good information about the wide impedance spectrum of different white cells. In other words, it is easy to differentiate between different white blood cells if their shapes are non-spherical shapes. The non-spherical shape can be produced by exposing the cells to a mechanical force before testing their impedance; for example by applying the testing cells to columb force via

applying an electric field to the cells.

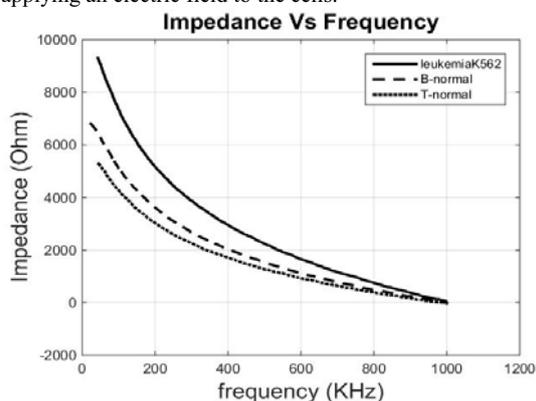


Fig. 11 The impedance of the normal and malignant cell of white blood cell as ellipsoid

c- Testing Considerations

In this section, the effect of the cell trapping between the sensing electrodes must be taken into account. The sensing capacitance value will be changed according to the cell position (i.e., near or on the electrodes). A maximum 10 % deviation is achieved (Fig. 11) in case the testing cells are located between the electrodes or part of these cells is located on the electrode surface by 50 μm (Fig.12).

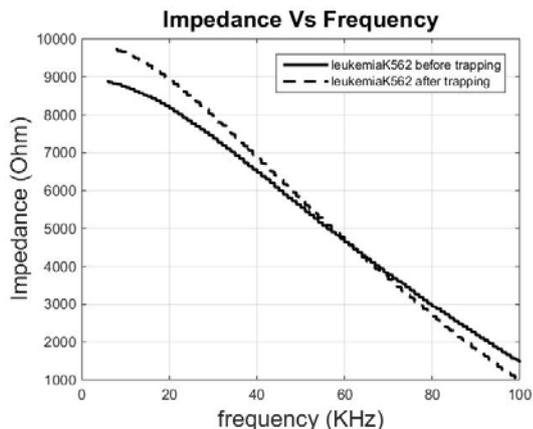


Fig. 12 The impedance of Leukemia before trapping and after trapping between electrodes

VII.. CONCLUSION AND REMARK

In this work, the Clausius-Mossotti factor (CM factor) of the normal and malignant white blood cells is studied analytically. Also, the effects of the single shell and double shell approximations are considered in this work. The Spherical and non-spherical geometries are considered in this work. An accurate 3d numerical simulation is introduced to extract the leukemia cell impedance. Also, the cell position and system design considerations are considered in this work.

ACKNOWLEDGEMENT

This research was partially funded by Zewail City of Science and Technology, AUC, the STDF, Intel, Mentor Graphics, ITIDA, SRC, ASRT and MCIT.

V. REFERENCES

- [1] F. Bordi, C. Cametti, A. Rosi, A. Calcabrini, "Frequency domain electrical conductivity measurements of the passive electrical properties of human lymphocytes", *Biochim. Biophys. Acta* 1153 (1993) 77-88
- [2] S. MacKay, P. Hermansen, D. Wishart and J. Chen, "Simulations of Interdigitated Electrode Interactions with GoldNanoparticles for Impedance-Based Biosensing Applications".
- [3] P. Gascoyne, X. Wag, Y. Huang, F. Becker, "Dielectrophoretic Separation of Cancer Cells from Blood", *IEEE Transactions on Industry Applications*, 33:3 (1997) 670-678.
- [4] H. Morgan, T. Sun, D. Holmes, S. Gawad and N. G. Green, "Single cell dielectric spectroscopy", *J. Phys. D: Appl. Phys.*, 40, pp. 61-70, (2007) *Cytochemistry*, 27: 1 (1979) 234-240.
- [5] S. Gawad, K. Cheung, U. Seger, A. Bertsch and P. Renaud, "Dielectric spectroscopy in a micromachined flow cytometer: theoretical and practical considerations", *Lab Chip*, 4, pp. 241-251, (2004)
- [6] H. Morgan, M. Hughes, N. Green, "Separation of Submicron Bioparticles by Dielectrophoresis", *Biophysical Journal*, 71 (1999) 516-525
- [7] A. Valero, T. Braschler, P. Renaud, "A unified approach to dielectric single cell analysis: Impedance and dielectrophoretic force spectroscopy". *Lab Chip* 2010, 10, 2216-2225.
- [8] F. Becker, X.-B. Wang, Y. Huang, R. Pethig, J. Vykoukal, P.R.C. Gascoyne, "Separation of human breast cancer cells from blood by differential dielectric", *Proc. Natl. Acad. Sci. USA* 92 (1995) 860-864.
- [9] A. Irimajiri, K. Asami, T. Ichinowatari, Y. Kinoshita, "Passive electrical properties of the membrane and cytoplasm of cultured rat basophil leukemia cells. I. Dielectric behavior of cell suspensions in 0.01-500 MHz and its simulation with a single-shell model", *Biochim. Biophys. Acta* 869 (1987) 203-213.
- [10] S. Gawad, M. Heuschkel, Y. Leung-Ki, R. Iuzzolin, L. Schild, Ph. Lerch, Ph. Renaud, "Fabrication of A Microfluidic Cell Analyzer In A Microchannel Using Impedance Spectroscopy".
- [11] M.T. Santini, C. Cametti, P.L. Indovina, G. Morelli, G. Donelli, "Polylysine induces changes in membrane electrical properties of K562 cells", *J. Biomed. Mater. Res.* 35 (1997) 165-174
- [12] L.M. Broche, F.H. Labeed, M.P. Hughes, "Extraction of dielectric properties of multiple populations from dielectrophoretic collection spectrum data". *Phys. Med. Biol.* 2005, 50, 2267-2274.
- [13] K. Asami, Y. Takahashi, S. Takashima, "Dielectric properties of mouse lymphocytes and erythrocytes", *Biochim. Biophys. Acta* 1010 (1989)
- [14] Rizzoni, Giorgio, "Principles and Applications of Electrical Engineering"
- [15] Y. Plevaya, I. Ermolina, M. Schlesinger, B.Z. Ginzburg, Y. Feldman, "Time domain dielectric spectroscopy study of human cells II. Normal and malignant white blood cells", *Biochimica et Biophysica Acta* 1419 (1999) 257-271
- [16] G. Fuhr and P. I. Kuzmin, "behaviour of cells in rotation electric fields with account to surface charges and cell structures", *Biophys. J.*, 50, pp. 789-795, 1986.
- [17] C. Bernabini, D. Holmes, H. Morgan, "Micro-impedance cytometry for detection and analysis of micron-sized particles and bacteria". *Lab Chip* 2011, 11, 407-412.
- [18] A. Ron, R. R. Singh, N. Fishelson, L. Shur, R. Socher, N. Croitoriu, D. Benayahu and Y. Shacham-Diamond, "Dielectric dispersion of suspended cells using 3D reconstructed morphology model", *Bioelectrochemistry*, 75(2)(2009)95-103.
- [19] M. Chattopadhyay, S. Sahu and D. Chowdhury "Determination of Malignancy in Leukocytes with Micro Capacitive Sensor", *International Journal of Biotechnology and Biomedical Sciences* p-ISSN 2454-4582, e-ISSN 2454-7808, Volume 2, Issue 2; January-June, 2016 pp. 130-133

Controlling drug delivery from multi-layer polymeric coated capsules

Giuseppe Pontrelli¹, Elliot J. Carr², Badr Kaoui³,
Marco Lauricella¹, and Sauro Succi¹

¹ Istituto per le Applicazioni del Calcolo - CNR,
Consiglio Nazionale delle Ricerche, Roma, Italy
{giuseppe.pontrelli@gmail.com}

² School of Mathematical Sciences,
Queensland University of Technology, Brisbane, Australia

³ Biomechanics and Bioengineering Laboratory (UMR CNRS 7338),
Sorbonne Universités, Université de Technologie de Compiègne, France

Nowadays there is a tremendous increasing interest in using capsules as targeted drug delivery systems. They allow enhanced therapeutic efficacy and reduce side effects by controlling the drug dose released in the human body. Capsules consist usually of a drug-loaded (fluid or solid) core surrounded by one or few hydrogel layers. Such encapsulation with multiple layers enhances the mechanical stability, biocompatibility, protects the active ingredients from external chemical aggression and premature degradation, and extends the sustainability of the drug release [1]. Different technologies have emerged in the last years to design and build layer-by-layer concentric spherical capsules [2], even though drug release rate cannot be easily predetermined. Depending on the nature of the encapsulated formulations and according to the final aimed therapeutic requirements, the typical size of capsules can range from some nano- to milli-meters. For biomedical applications, micro-capsules are largely used [3].

For some specific applications, a thin coating shell is required to envelop the whole capsule structure in order to protect it from external chemical aggressions and mechanical erosion. Drug release characterization consists in tracking the kinetics of the drug eluted from the capsule into the external targeted medium, which is away from being an easy task. However, effort and costs of developing and designing new optimized delivery devices can be dramatically reduced if the release mechanism is understood in advance using appropriate computational models. Here, we are interested in mechanistic models, because they allow a better understanding of the underlying mechanism of the drug release by tracking the influence of each physical input parameter, in contrast with the empirical approach. We upgrade existing mechanistic models by extending their applications to the composite coated capsules. Thus, predictable simulations based on those models can spot the significant physical parameters and allow extracting reliable information for conducting in vitro experiments. Diffusion is by far the domi-

nant mechanism in drug delivery, beside other physico-chemical factors, such as osmosis, drug dissolution, and polymer swelling or degradation.

In this work, we will make a step forward by studying theoretically and computationally the drug release from a multi-layer microcapsule, under the assumption of radial symmetry, via an analytic procedure, and moreover by avoiding considering simplistic hypothesis made elsewhere previously, such as a continuous concentration across the surface. We present a pure diffusive model through a composite spherical core-layer capsule coated with a protective thin shell. The problem of release in such a heterogeneous layer-by-layer composite medium is described by a system of coupled partial differential equations, which we solve analytically in terms of Fourier series [4] or using other exact closed-form solutions [5]. In addition to the conventional partitioning and mass transfer interlayer conditions, we consider a finite mass transfer resistance, to model the coating at the capsule surface (fig. 1).

Analytic expressions for the concentration and the cumulative mass in all layers are given to study the release system to show the dependence and sensitivity to parameters. The drug concentration profiles in the core and through all the layers, as well as in the external release medium, describe and characterize the drug release mechanism [4,5].

Preliminary results show that the drug mass is monotonically decreasing in the core, but is first increasing up to some upper bound and then decaying asymptotically in the hydrogel layer. In the release medium the mass progressively accumulates over an extended distance at a time depending on the diffusive properties of the two-layer capsule. The simulation points out the time and the size of the mass peak in the intermediate layer is related to the releasing properties of the core at one hand, and to the diffusivity of the release medium at the other hand, together with the mass resistance of the coating. It appears also that the relative size of the layers and their respective diffusivity affect the whole drug release processes. Depending on the specific application, it is worth identifying which set of parameters guarantees a more prolonged and uniform release and what other values are responsible for a localized peaked distribution followed by a faster decay. One of them is the permeability of the coating shell that offers a significant resistance to the mass flux. Thus, the coating parameter has to be properly tuned in order to allow drug molecules to be released, while maintained in the efficient therapeutic range without exceeding the toxic dose nor dropping below an insufficient dose. These results can be used to assess whether drug targets tissues at the desired rate and to optimize the dose capacity given by thin surface coating shells for an extended period of time. Differently than in other single layer models, the current formulation constitutes a simple tool to predict the accurate release from a multi-layer coated capsule that can help in designing and in manufacturing new drug delivery platforms.

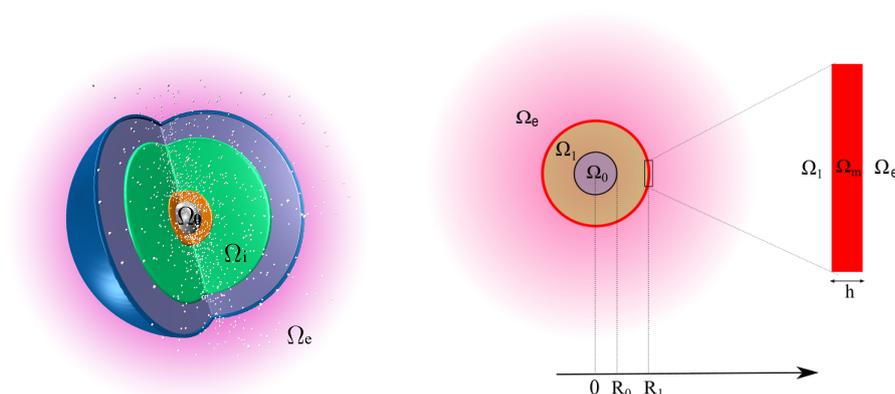


Fig. 1. At left, drug releasing from a multi-layer microcapsule. Drug is initially loaded in the core Ω_0 and diffuses, through all the intermediate layers, into the release medium Ω_e . Suitable conditions are set at internal interlayer interfaces and at the external coating shell. At right, a schematic representation of the cross-section of the radially symmetric two-layer capsule, made of a central core Ω_0 , a concentric shell Ω_1 and a thin protecting coating Ω_m (in red). Together with the external release medium Ω_e , it constitutes a three concentric layers system. On the right side, a zoom of the coating layer (figure not to scale).

References

1. De Koker S., Hoogenboom R., De Geest B.G., Polymeric multilayer capsules for drug delivery, *Chem. Soc. Rev.*, 41, 2867–2884 (2012)
2. Timin A.S., Gould D.J., Sukhorukov G.B., Multi-layer microcapsules: fresh insights and new applications, *Exp. Opin. Drug. Del.*, 14(5) 583–587 (2016)
3. Larranaga A., Lomora M., Sarasua J.R., Palivan C.G., Pandit A., Polymer capsules as micro-/nanoreactors for therapeutic applications: Current strategies to control membrane permeability, *Prog. Mater. Science*, 90, 325–257 (2017)
4. Kaoui B., Lauricella M., Pontrelli G., Mechanistic modelling of drug release from multi-layer capsules, *Comp. Biol. & Med.* 93, 149–157 (2018)
5. Carr E., Pontrelli G., Modelling mass diffusion for a multi-layer sphere immersed in a semi-infinite medium: application to drug delivery, submitted (2018)

Drug diffusion properties of hydrogels: numerical and experimental studies

S. Solodova¹, A.F.R. Pimenta², K. Gudnason³, A.P. Serro², F. Jonsdottir³, S. Th. Sigurdsson³, M. Masson¹

¹ Faculty of Pharmaceutical Science, University of Iceland, Hofsvallagata 53, IS-107 Reykjavik, Iceland

² University of Lisbon, Av. Rovisco Pais, 1049-001 Lisbon, Portugal

³ Faculty of Industrial Engineering, Mechanical Engineering and Computer Science, University of Iceland, Hjardarhagi 2-6, IS-107 Reykjavik, Iceland
Svetlana@hi.is

Abstract. Hydrogel materials for drug delivery systems are very extensive field of scientific explorations due to their unique properties. Understanding the drug delivery properties is key question to the designing of therapeutic drug delivery systems. In this study, we present a general one dimensional multi-layer model based on experimental investigation of diffusion properties for two types hydrogels (pHEMA and TRIS) at different temperatures. Model can be applied to simulate results of diffusion experiments but also applies to more complex simulations.

Keywords: drug delivery, hydrogels, numerical modelling.

Hydrogel materials are highly promising polymeric biomaterials which attract the attention of a large number of researchers of different areas, such as food industry, pharmaceutical, biomedical, drug delivery system, etc., due to their important properties like hydrophilicity, biocompatibility, good transport properties and non-toxicity. Hydrogels can also be made biodegradable or bioabsorbable for specific applications. More developments are expected in drug delivery and tissue engineering [1-3]. Currently, due to significant advances in computer simulation technology, numerical modeling is increasingly becoming an essential part of experimental studies. The mathematical description of mass transport processes such as diffusion transport through different drug delivery systems, drug dissolution, drug release and other can be very helpful to better understanding the mechanisms of such phenomena and consequently to help develop optimization strategies for targeting drug delivery [4]. The present study was devoted to experimental and numerical studies of transport properties of two types of polymer materials used in the production of contact lenses: poly(2-hydroxyethyl methacrylate) (pHEMA) based hydrogel and 3-tris(trimethylsilyloxy)silylpropyl 2-methylprop-2-enoate (TRIS) silicone based hydrogel. Diffusion properties of model drug diclofenac across hydrogels were investigated by in vitro diffusion study (Franz diffusion cells experiment) at different temperatures (25, 35, 45 °C). The lag-time method has proven to be an effective method for

characterization permeation processes, because of the simple nature of the permeation experiment, transport parameters can be directly obtained from experimental data. We present a general one dimensional multi-layer model that can be applied to simulate results of Franz diffusion cell experiments but also applies to more complex simulations [5].

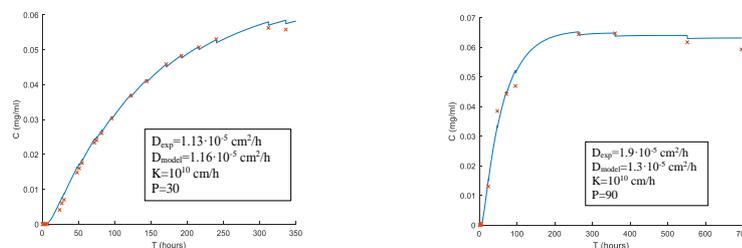


Fig.1. Simulated drug diffusion curves for pHEMA at T=25 °C and T=35 °C, correspondingly (× experiment, – modelling).

The physical parameters for hydrogels were estimated from drug diffusion experiments at different temperatures. Effects and interdependencies of physical parameters on the profiles of pHEMA and TRIS at different temperatures were analyzed to see which part of the curve is influenced by which parameter. We succeeded in simulating all experimental data sets.

References

1. Nagam, S.P., Jyothi, A.N., Poojitha, J., Aruna, S., Nadendla, R.R.: A comprehensive review on hydrogels. *International Journal of Current Pharmaceutical Research* 8(1), 19-23 (2016).
2. Hoffman, A.S.: Hydrogels for biomedical applications. *Advanced Drug Delivery Reviews* 64, 18-23 (2012).
3. Rowley, J., Madlambayan, G., Faulkner, J., Mooney, D.J.: Alginate hydrogels as synthetic extracellular matrix materials. *Biomaterials* 20, 45-53 (1999).
4. Bozsak, F., Gonzalez-Rodriguez, D., Sternberger, Z., Belitz, P., Bewley, T., Chomaz, J.-M., Barakat, A.I.: Optimization of Drug Delivery by Drug-Eluting Stents. *PLoS ONE* 10(6), 1-29 (2015).
5. Gudnason, K., Solodova, S., Vilardell, A., Masson, M., Sigurdsson, S., Jonsdottir F.: Numerical simulation of Franz diffusion experiment: Application to drug loaded soft contact lenses. *Journal of Drug Delivery Science and Technology* 38, 18-27 (2017).

Numerical evaluation of skin sub layer properties

Bergthóra S. Snorradóttir¹, Kristinn Guðnason², Fjóla Jónsdóttir², Sven Th. Sigurdsson² and Már Másson¹

¹ Faculty of Pharmaceutical Science, University of Iceland, Hofsvallagata 53, 107 Reykjavík, Iceland

² Faculty of Industrial Engineering, Mechanical Engineering and Computer Science, University of Iceland, Hjardarhagi 2-6, 107 Reykjavík, Iceland

lncs@springer.com

Abstract. Modelling of transdermal drug delivery has been of much interest recently, since transdermal absorption is a complicated mechanism affected by number of factors. Adhesive tape stripping measurement and confocal microscopy have been used to determine which the dominant route of penetration is. The purpose of *In vitro* studies to determine transdermal permeation, penetration and absorption can vary and thus different approaches and models are used for such studies. Franz diffusion cell are often chosen as tools for *in vitro* experiments. There are several distinct advantages of using in vitro diffusion cells to study transdermal drug penetration. An important advantage over in vivo studies is that skin layers can be separated and studied in isolation. A diffusion based skin model should include the barrier effects of the stratum corneum as well as partitioning. This work considers modelling of the transdermal delivery route .

Keywords: Transdermal delivery, Numerical modelling, Franz diffusion cell.

Modelling of transdermal penetration

Transdermal drug delivery is recognized as a successful and feasible route of absorption despite the challenges of moving drugs across the skin. The purpose of *In vitro* studies to determine transdermal permeation, penetration and absorption can vary and thus different approaches and models are used for such studies. Franz diffusion cell (fig.1) are often chosen as tools for *in vitro* experiments [1]. There are several distinct advantages of using in vitro diffusion cells to study transdermal drug penetration. An important advantage over in vivo studies is that skin layers can be separated and studied in isolation (e.g., heat separation or tape stripping to remove layers of stratum corneum (SC)).

The full thickness skin includes SC, epidermis and dermis and using Frans diffusion cell experiments allows characterization of drug penetration that would not be possible in a living animal, including layers of the SC and the less lipophilic regions of the lower epidermis and dermis.

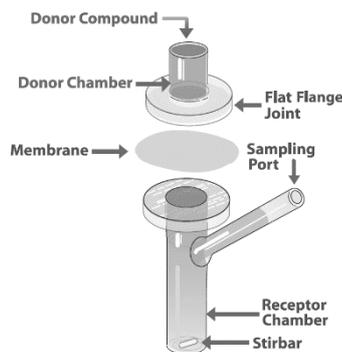


Fig. 1. Schematic figure of a Franz diffusion cell.

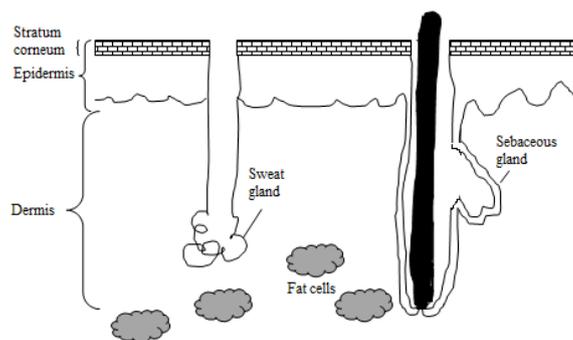


Fig. 2. Skin structure of full thickness skin including stratum corneum, epidermis and dermis.

Different drug formulations have different properties within the skin sublayers (SC, epidermis and dermis) and need to be understood in order to construct effective regimen. By applying fluorescent chemicals or using fluorescent drug formulation it is possible to observe the concentration profile within particular skin layers using confocal microscopy. This provides valuable insight when evaluating drug properties within the layers as it allows for comparison to simulated concentration directly.

Modelling of transdermal drug delivery is very interesting, it is important that the model is simple enough for scientists to use as well as being able to explain complex data [2], since transdermal absorption is a complicated mechanism affected by number of factors. Therefore there is great interest in developing a model that can rather accurately predict transdermal drug transport including the transport through the skin. A diffusion based skin model should include the barrier effects of the stratum corneum as well as partitioning [3] to account for the difference in lipophilicity between layers. The observable concentration profile aids in assessing the lipophilicity.

References

1. Snorradóttir, B.S., Gudnason, P.I., Thorsteinsson, F., Másson, M., 2011. Experimental design for optimizing drug release from silicone elastomer matrix and investigation of transdermal drug delivery. *European Journal of Pharmaceutical Sciences*, 42, pp.559-567.
2. Snorradóttir, B.S., Jonsdóttir, F., Sigurdsson, S. and Masson, M., 2014. Numerical modeling of transdermal delivery from matrix systems: parametric study and experimental validation with silicone matrices. *Mathematical biosciences Journal of Pharmaceutical Sciences*, 103, pp.2366-2375.
3. Gudnason, K., Sigurdsson, S., Snorradóttir, B.S., Masson, M. and Jonsdóttir, F., 2018. A numerical framework for drug transport in a multi-layer system with discontinuous inter-layer condition. *Mathematical biosciences*, 295, pp.11-23.

Modelling the drug release from intraocular lens material with variable loading times

Ana Topete ^[1], Kristinn Gudnason ^[2], B. Saramago ^[1] and A. P. Serro ^[3]

¹ Centro de Química Estrutural, Departamento de Engenharia Química, Instituto Superior Técnico, Universidade de Lisboa, Av. Rovisco Pais, 1049-001 Lisboa, Portugal

² Faculty of Industrial Engineering, Mechanical Engineering and Computer Science, University of Iceland, Iceland

³ Centro de Investigação Interdisciplinar Egas Moniz, Instituto Superior de Ciências da Saúde Egas Moniz, Quinta da Granja, Monte de Caparica, 2829-511 Caparica, Portugal

anatopete@tecnico.ulisboa.pt

Cataracts are the main cause of blindness worldwide. Their treatment involves surgery, in which the cloudy natural lens of the eye is removed and an intraocular lens (IOL) is implanted in its stead. After surgery, antibiotics and anti-inflammatories are prescribed in the form of eye drops, with a high frequency (until 4x daily) for long periods of time (at least 2 weeks for antibiotics and 4 weeks for anti-inflammatories), to prevent complications such as endophthalmitis. This drug delivery method is not very effective since it depends on the patient compliance and more than 95% of the drug is lost due to protective eye mechanisms and systemic absorption. Drug-loaded IOLs are a promising alternative to overcome these problems. However, a burst in the beginning of the release is usually experienced and the release kinetics is not adequate. The objective of the work is to optimize the loading conditions (time and temperature) of IOLs to achieve a sustained drug delivery of diclofenac and moxifloxacin. Mathematical modelling was applied to gain more insight into the drug related properties of the IOL material, which are essential to predict the drug release behavior.

Discs of CI26Y (Contamac UK), a copolymer of hydroxyethyl methacrylate (HEMA) and methyl methacrylate (MMA), used to produce IOLs were loaded with moxifloxacin hydrochloride and diclofenac sodium salt by soaking in drug solution (5 mg/mL) at 4°C, for 4 days, 2 weeks and 1, 2 and 3 months and, at 60°C, for 4 days and 1 and 2 weeks. Drug release experiments were carried out in closed tubes containing PBS, at 36°C, with stirring. The concentration of the drug released was determined by UV-Vis spectroscopy. To model these experiments, various processes need to be taken into account, such as the diffusion coefficient within the lens, the partition coefficient, and the mass transfer coefficient, describing possible surface barrier effects. To this end, the following set of equations were employed:

$$\frac{\partial C_{\alpha}(x, t)}{\partial t} = \nabla \cdot (D \nabla C_{\alpha}(x, t)) \quad x \in \Omega$$

Equation 1

$$-D_{\alpha}(\nabla C_{\alpha}(x, t) \cdot \mathbf{n}) = K(C_{\alpha}(x, t) - PC_{\beta}(t)) \quad x \in \Gamma$$

Equation 2

The equation 1 describes diffusion within the lens and the equation 2 describes the general boundary conditions. Ω refers to the region of an IOL disc and Γ refers to its boundary. C_{α} and C_{β} are concentration distributions within the lens and in the PBS respectively. D , P , K refer to the diffusion coefficient within the lens, partition coefficient, and mass transfer resistance respectively.

For moxifloxacin, at both temperatures, higher loading time increased the drug release, being this difference more notorious at 4°C between 2 weeks and 1 month. The higher moxifloxacin release was achieved with the higher loading temperature (60°C) after 2 weeks of loading. For diclofenac, a different behavior was seen. At 4°C higher loading time increased the drug release, until the equilibrium was achieved after 2 months of loading. At 60°C, different loading time did not affect the amount of drug released. The higher diclofenac release was achieved at 4°C, after 2 months of loading. The experiments were simulated and compared to the release data to evaluate temperature sensitive drug-material specific coefficients.

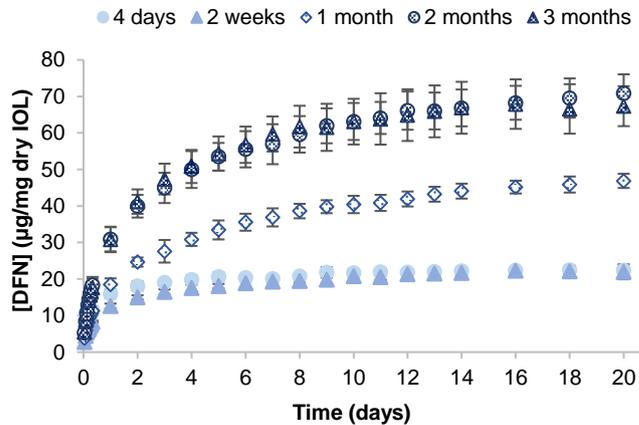


Figure 1. Cumulative release profiles of diclofenac from CI26Y hydrogels loaded at 4 °C for different periods of time.

The prognostic value of clinical variables for the treatment response in rheumatoid arthritis

Alexander Platzer¹, Daniela Sieghart¹, Farideh Alasti¹, Guenter Steiner¹, Josef Smolen¹

¹ Division of Rheumatology, Department of Internal Medicine III, Medical University of Vienna, Vienna, Austria

Keywords: rheumatoid arthritis, clinical variables, autoantibodies, rheumatoid factor, anti-citrullinated protein antibodies, methotrexate, treatment response, treatment outcome, therapy response prediction, machine learning

Abstract. Several therapies are available for treating rheumatoid arthritis (RA), but the final success rate of a single therapy is limited. In most of the cases, different therapies are applied consecutively. The main recommendations for the treatment of RA are from the American College of Rheumatology (ACR [1]) and from the European League Against Rheumatism (EULAR [2]). The outcome or response to RA-therapies is not a single severe effect, but a more long-term worsening or improvement. We took the measures of the two main recommendations: SDAI50 (simplified disease activity score [3]) and ACR20 [4]. The 1995 developed ACR response criteria comprises seven core set variables: swollen joint count (SJC), tender joint count (TJC), physician's assessment of disease activity (PGA), patient's assessment of pain, and patient's assessment of physical function, and levels of an acute-phase reactant - either C-reactive protein (CRP) or erythrocyte sedimentation rate (ESR).

In addition, the simplified disease activity index (SDAI) is the numerical sum of TJC and SJC (based on a 28-joint assessment), PGA, EGA and CRP (mg/dl, normal <1 mg/dl). 50% improvement in SDAI (SDAI50) was defined as minor response [5] - equivalent to ACR20. ACR20 or SDAI50 are used as therapy response variables in this analysis.

In the two main recommendations are also sections on prognostic markers, but the consensus so far is far from complete. Most of the markers therein were originally investigated on a per-single-marker basis; for several other markers are reports published, but their prognostic value has still not been fully elucidated.

Here, we investigate about 80 clinical variables, whose data is present for most of about 2000 patients in our main database. Additionally, we report some insights of additional clinical variables, which were measured for studies of smaller sets of patients. The focus of this study is on models, not just single variables (as e.g. reviewed in Cuppen et al. [6]).

The clinical variables consist of measurements for RA (like SJC28, HAQ (Health Assessment Questionnaire), PGA), autoantibodies (RF and CCP) and

the hemogram (KREA (creatinine), ALAT (alanine aminotransferase), BUN (blood urea nitrogen), etc.).

The about 2000 patients are Austrian with diagnosed RA. The distribution of patient characteristics follows the expectation at RA [7]: the majority is female and older than 55 years.

The first treatment for a RA patient is usually methotrexate (MTX), following the recommendations of ACR and EULAR. Therefore, most of the patients were treated with MTX. After stringent filtering (data for most clinical variables present at start of the treatment and present data at the time an effect is expected, that is after 2.5 to 7.5 months), we have 151 patients left.

We applied several machine learning classification methods: SVMs, tree-learners, naive Bayes, boosting and some more implemented in weka ([8]) to generate prediction models. One of the best performing and also one of the simplest model is the generated alternating decision tree (which is significant and has a cross-fold accuracy of 75% and 93% on training data; for general information on alternating trees see the paper of Freund et al. [9]).

The most important variables - according to the information gain - are related to the severity of the disease in the beginning and the status of the immune system.

Another class of treatments are those targeting TNF, called anti-TNF treatments. Here, the picture was not so clear.

From the diagnostic side, anti-citrullinated protein antibodies (ACPA) and rheumatoid factor (RF) are the most specific diagnostic markers of rheumatoid arthritis [10]. These and other subtypes and other antibodies have been reported (e.g. [11]), but their prognostic value is not fully described. We investigated the IgA, IgG and IgM subtypes of RF, ACPA and RA33. For our data we found a significant alternating decision tree, with a cross-fold accuracy of 74% (accuracy on training data 77%). This alternating decision tree can be even shortened to a simple decision tree.

Unfortunately, there is only a small overlap between the patients from which we measured the extended set of antibodies and the patients from which we have the more complete general clinical data, so we can so far only roughly compare the models of these two datasets.

It is planned for the future to have these and more variables consistently measured in a bigger group of patients. Beside antibodies against different targets, also the subclasses might be interesting (IgA1, IgA2, IgG1-IgG4). It remains a challenge to generate the data, but also to convince different studies/colleagues to make their data ready for combining to a bigger set.

1. Singh, J.A., Saag, K.G., Bridges, S.L., Jr., Akl, E.A., Bannuru, R.R., Sullivan, M.C., Vaysbrot, E., McNaughton, C., Osani, M., Shmerling, R.H., Curtis, J.R., Furst, D.E., Parks, D., Kavanaugh, A., O'Dell, J., King, C., Leong, A., Matteson, E.L., Schousboe, J.T., Drevlow, B., Ginsberg, S., Grober, J., St Clair, E.W., Tindall, E., Miller, A.S., McAlindon, T., American College of, R.: 2015 American College of Rheumatology Guideline for the Treatment of Rheumatoid Arthritis. *Arthritis Care Res (Hoboken)* 68, 1-25 (2016)

2. Smolen, J.S., Landewe, R., Bijlsma, J., Burmester, G., Chatzidionysiou, K., Dougados, M., Nam, J., Ramiro, S., Voshaar, M., van Vollenhoven, R., Aletaha, D., Aringer, M., Boers, M., Buckley, C.D., Buttgereit, F., Bykerk, V., Cardiel, M., Combe, B., Cutolo, M., van Eijk-Hustings, Y., Emery, P., Finckh, A., Gabay, C., Gomez-Reino, J., Gossec, L., Gottenberg, J.E., Hazes, J.M.W., Huizinga, T., Jani, M., Karateev, D., Kouloumas, M., Kvien, T., Li, Z., Mariette, X., McInnes, I., Mysler, E., Nash, P., Pavelka, K., Poor, G., Richez, C., van Riel, P., Rubbert-Roth, A., Saag, K., da Silva, J., Stamm, T., Takeuchi, T., Westhovens, R., de Wit, M., van der Heijde, D.: EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2016 update. *Ann Rheum Dis* 76, 960-977 (2017)
3. Smolen, J., Breedveld, F., Schiff, M., Kalden, J., Emery, P., Eberl, G., Van Riel, P., Tugwell, P.: A simplified disease activity index for rheumatoid arthritis for use in clinical practice. *Rheumatology* 42, 244-257 (2003)
4. Felson, D.T., Anderson, J.J., Boers, M., Bombardier, C., Furst, D., Goldsmith, C., Katz, L.M., Lightfoot, R., Jr., Paulus, H., Strand, V., et al.: American College of Rheumatology. Preliminary definition of improvement in rheumatoid arthritis. *Arthritis Rheum* 38, 727-735 (1995)
5. Aletaha, D., Martinez-Avila, J., Kvien, T.K., Smolen, J.S.: Definition of treatment response in rheumatoid arthritis based on the simplified and the clinical disease activity index. *Annals of the rheumatic diseases* 71, 1190-1196 (2012)
6. Cuppen, B.V., Welsing, P.M., Sprengers, J.J., Bijlsma, J.W., Marijnissen, A.C., van Laar, J.M., Lafeber, F.P., Nair, S.C.: Personalized biological treatment for rheumatoid arthritis: a systematic review with a focus on clinical applicability. *Rheumatology (Oxford)* 55, 826-839 (2016)
7. Curtis, J.R., Jain, A., Askling, J., Bridges, S.L., Jr., Carmona, L., Dixon, W., Finckh, A., Hyrich, K., Greenberg, J.D., Kremer, J., Listing, J., Michaud, K., Mikuls, T., Shadick, N., Solomon, D.H., Weinblatt, M.E., Wolfe, F., Zink, A.: A comparison of patient characteristics and outcomes in selected European and U.S. rheumatoid arthritis registries. *Semin Arthritis Rheum* 40, 2-14 e11 (2010)
8. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter* 11, 10-18 (2009)
9. Freund, Y., Mason, L.: The alternating decision tree learning algorithm. In: *icml*, pp. 124-133. (Year)
10. Aho, K., Heliovaara, M., Maatela, J., Tuomi, T., Palosuo, T.: Rheumatoid factors antedating clinical rheumatoid arthritis. *J Rheumatol* 18, 1282-1284 (1991)
11. Goldbach-Mansky, R., Lee, J., McCoy, A., Hoxworth, J., Yarboro, C., Smolen, J.S., Steiner, G., Rosen, A., Zhang, C., Menard, H.A., Zhou, Z.J., Palosuo, T., Van Venrooij, W.J., Wilder, R.L., Klippel, J.H., Schumacher, H.R., Jr., El-Gabalawy, H.S.: Rheumatoid arthritis associated autoantibodies in patients with synovitis of recent onset. *Arthritis Res* 2, 236-243 (2000)

The Study of Complications of Intestinal Obstruction Catheter

BIN SONG¹, HAI-XIA HU¹, JUAN DU², and YAN-ZHUO SHU¹

¹Department of Gastrointestinal, Colorectal and Anal Surgery, China-Japan Union Hospital of Jilin University, Changchun, Jilin 130033; ²Internal Medicine 2, The Tumor Hospital of Jilin Changchun, Jilin 130012, P.R. China

Objective: To probe the severe complications in the application of intestinal obstruction catheter.

Methods: We retrospectively analyzed the clinical data of 142 patients who were treated by intestinal obstruction catheter in the gastrointestinal surgery department of China-Japan Union Hospital of Jilin University from December 2006 to September 2017, and analyzed the serious complications.

Results: There are 12 cases of severe complications in this group, and the incidence is 8.45%. 1 case of shedding of fore-end of the catheter from the anus; 4 cases of rupture of anterior balloon; 3 cases of intestine intussusception after extubation; 1 case of knotting inside the gastric cavity; 1 case of catheter hydraulic piston obstruction, and 2 cases of catheter obstruction and cannot be dredged.

Conclusion: The serious complications of the intestinal obstruction catheter should be paid attention to in clinical application, timely treatment after onset of symptoms and attach importance to standard operation and active prevention.

[Key words] Intestinal Obstruction Catheter; Complication

Acute intestinal obstruction is one of the most common acute abdomen diseases, which accounts for 20%^[1] of the incidence of acute abdomen diseases. Except for the strangulated obstruction needs emergency surgery, the other obstructions all can be treated with expectant treatment, gastrointestinal decompression is the main measure to alleviate intestinal pressure. Intestinal obstruction catheter is a new method to alleviate intestinal pressure, which can effectively and rapidly remove the obstruction symptoms^[2]; for the complex and recurrent intestinal obstruction it is possible to adopt intestinal obstruction catheter plication of small intestine. During the process of massive application of intestinal obstruction catheter, our department summarized the serious complications of intestinal obstruction catheter and made reasonable prevention and treatment. The detailed contents of the report are as follows:

1. Data and Methods

1.1 General Information

The study objects are 142 cases of patients who were treated by via-nose intestinal obstruction catheter in the gastrointestinal surgery department of China-Japan Union Hospital of Jilin University from December 2006 to September 2017 (including 68 cases of via-nose intestinal obstruction catheter plication of small intestine), observing the relevant complications of intestinal obstruction catheter in the treatment process. The selection of case does not include the following cases: strangulated intestinal obstruction, severe cardiopulmonary insufficiency, high obstruction, patients do not compliance to treatment. All patients all firstly remain the gastric intubation

after admission, and the gastrointestinal decompression was not relieved obviously. Among them, there are 72 cases of men and 70 cases of women; aged between 38 -83 years old, with an average age of 54.4 years. The orthostatic X-ray abdominal plain film and abdominal CT examination were performed before operation. The adopted intestinal obstruction catheter is three-cavity two- balloon (front balloon, rear balloon) via-nose intestinal obstruction catheter produced by Dalian Kuliaite Medical Treatment Products Co.Ltd.

1.2 Operation Method

The intestinal obstruction catheter is indwelled to the stomach via nose, and guide the catheter to the descending section of duodenum under the guidance of gastroscopy, to ensure the second balloon is delivered to the pylorus; infuse 10-15ml distilled water to the front balloon of the catheter, and pull out the hydrophilic guide wire and confirm its position by X-ray, and reconfirm the intestinal obstruction catheter is in descending section of the duodenum. The catheter will move to obstruction as the peristalsis. After conservative treatment, 74 cases of patients gradually discharged and defecated and the intestinal obstruction were relieved.

1.3 Operation Methods

1.3.1 Surgical Treatment of Adhesive Intestinal Obstruction

54 cases of adhesive intestinal obstruction were treated with conservative treatment, although with relief of abdominal distension, but no exhaust and defecation. Laparotomy was performed to remove adhesion, at the same time the intestinal obstruction catheter plication of small intestine was applied. The preoperative indwelling catheters of 32 cases were guided to ileocecus by hand. The retrograde plication was performed for 12 cases, namely the retrograde catheterization was arranged in the small intestine from ileocecus. The catheter of 10 cases was pulled out via nose and gastrostomy intubation was performed for place.

1.3.2 Surgical Treatment of Malignant Intestinal Obstruction

14 cases of malignant intestinal obstruction caused by extensive metastasis of tumor were relieved by conservative treatment. The intestinal obstruction catheter was pulled out through the nose, the gastrostomy intubation was performed and there was no obvious stenosis and the 10 ml water balloon of intestinal obstruction can pass more than 2m long small intestine, meanwhile, the enterostomy was performed.

1.4 Extubation Time

Non-surgical patients, after exhausting and defecation, then fix the catheter and pump out the balloon, gradually feed liquid diet until semi-liquid diet; if the patient still without any discomfort, then gradually pull out catheter; as for patients of adhesive intestinal construction anterograde plication of intestine, one week after operation, if the radiography is unobstructed, then gradually pull out the catheter; as for patients of adhesive intestinal construction retrograde plication of intestine, two weeks after operation, gradually pull out the catheter; as for the malignant intestinal obstruction, long time indwelling catheter, do not pull it out if without obvious complications.

2. Results

There were 12 cases of severe complications in this group, and the incidence was 8.45%. 1 case of shedding of fore-end of the catheter from the anus, the incidence rate was 0.70% (1/142) ; 4 cases of rupture of anterior balloon, the incidence of

2.81%(4/142); 3 cases of intestine intussusception after extubation, the incidence of 2.11%(3/142); 1 case of knot inside the catheter gastral cavity, the incidence rate was 0.70% (1/142) ; 1 case of catheter water balloon obstruction the incidence rate was 0.70% (1/142) , and 2 cases of catheter obstruction, unable to perform negative pressure suction and the guide wire cannot be dredged, the incidence rate was 1.40% (2/142) .

3 Discussions

Gastrointestinal decompression is the most commonly used and basic conservative treatment for intestinal obstruction [3], the traditional nasogastric tube decompression can only drain the fluid in the gastric cavity, while the fluid in the intestinal cavity cannot be drained adequately [4]. The via-nose intestinal obstruction catheter can make full use of the intestinal peristalsis and move slowly in the intestine. It can effectively drain for the above part of the obstruction and relieve or remove the intestinal obstruction [5]. At present, it is widely used in clinic. Although the intestinal obstruction catheter has high safety, there are still some complications during the application, some complications even need surgical intervention, which requires doctors to be vigilant and perform reasonable treatments. No massive hemorrhage and intestinal perforation were found in this group, and all the complications were relieved after reasonable treatments.

The knot inside the catheter gastral cavity is rare, and the incidence of this group is 0.70%. The distal end is blocked and cannot directly pull out the catheter via the nose. intestinal obstruction catheter is curled and swirled in gastric cavity, indwelling is relatively common, but the knot inside the stomach is rare, mainly because of the slow progress of the distal end of the catheter, and too fast and too long insertion of near end via nose, it will be gradually relieved as the gradual progress of the distal end. We reinserted the guide wire for support, gradually adjust and relieve the knot by gastric forceps under gastroscopy. As for the aspect of prevention, Xu Xiaolu et al [6] think that after the implantation of catheter, the patient should receive X-ray examination 1-2 times in each day, to confirm the catheter position and make timely adjustment; at the same time we believe that we should avoid artificial insertion of the catheter, the progress of the catheter is based on the gravity, and it is possible to take a small amount of paraffin oil by oral for the purpose of lubrication.

Catheter blockage is relatively common; it can be relieved by washing or guiding wire dredging. But 2 cases in this group cannot be dredged because of the large particles of food residue, therefore we pulled out and replaced the catheter .When intestinal contents are undigested food residues or thicker stools, if the drainage volume is low or no liquid is elicited, it is possible to inject the 20 mL normal saline by injector for douching of catheter, so as to prevent obstruction [7] douching several times a day to avoid obstruction.

Intussusception is a serious complication of extubation, with an incidence of 2.11% in this group, 2 cases of anterograde plication of intestine; 1 case of indwelling intestine obstruction catheter via nose. 3 cases were cured by emergency operation. Pay attention to that extubation should not be pulled out in one time, but should be

completed in 3 to 5 days and 2-3 times in each day, gradually pull it out; while extubation, the patient should be given paraffin oil by oral, and then slowly and uniformly pull it out, do not pull it under strong negative pressure [8]; patients should eat immediately after each extubation so as to promote intestinal peristalsis and avoid intussusception.

The incidence of rupture of catheter balloon is 2.81% in this group, the catheter loss of action and cannot move along with the intestinal peristalsis, it should be replaced by new catheter and indwelling again. As for the treatment process try to avoid excess injection of liquid into the balloon, generally not more than 20 milliliters, so as to avoid the balloon rupture, at the same time intestinal arrangement during the operation should be gentle.

There is 1 case of balloon obstruction of catheter in this group caused by the injection of normal saline in the balloon, and cannot be pulled out of the throat during extubation and cannot be pulled out directly, then the balloon back to the stomach, and the balloon was punctured under gastroscopy. The treatment process should pay attention to that does not inject saline into the balloon so as to avoid the obstruction caused by crystallization [9].

The case of defecation of catheter via the anus is rare in clinic, and the incidence rate of this group is 0.85%. This case has received subtotal colectomy due to constipation. The catheter completely entered into the nasal cavity, and the front end was defecated through the anus. The patient had severe colic in the abdomen, and the symptoms were relieved by cutting the front end of the catheter and pull out via the nose.

In short, the clinical application of via-nose intestinal obstruction catheter is safe and effective, worthy of extensive clinical application, but there are still rare and severe complications, need to strictly enforce the rules of operation in the process of treatment, carefully observation of drainage, regularly recheck imageological examination and adjust at any time, so as to minimize the occurrence of serious complications.

Reference

- [1] Tian Chunjiang, Xu Xiang Feng, Li Guohua. The Application of Small Intestine Radiography via Intestinal Obstruction Catheter in the Diagnosis of Intestinal Obstruction [J]. Journal of Clinical Radiology, 2014, 33 (10): 1600-1603.
- [2] Tian Chunjiang, Li Guohua. The Application Progress of Via-Nose Type Intestinal Obstruction Catheter [J]. Chinese Journal of Integrated Chinese and Western Medicine, 2014, 20 (5): 571-573.
- [3] Yang Yang, Li Liangqing, Chen Qun. Treatment of Non-strangulated Intestinal Obstruction by Via-Nose Intestinal Obstruction Catheter [J]. Guangdong Medical, 2013, 34 (10): 1573-1574.

- [4] Hu Liang, Ye Sheng, Ye Long Yun, et al. Report of 2 Cases of Treatment for Acute Intestinal Obstruction by Intestinal Obstruction Catheter [J]. Chinese Journal of Practical Surgery, 2012, 32 (11):966-968.
- [5] Wang Zhong, Li Bin, Han Tingchao, et al. Clinical Observation on the Treatment of Adhesive Intestinal Obstruction by Via-Nose Intestinal Obstruction Catheter. [J]. Clinical Misdiagnosis and Mistreatment, 2011, 24 (9): 54-56.
- [6] Xu Xiaolu, Chen Yongqiang, Gang group, et al. Clinical Application of Treatment of Intestinal Obstruction by Via-Nose Intestinal Obstruction Catheter under X-ray [J]. Chinese General Medicine, 2013, 11 (5): 790-792.
- [7] Sun Jun, Li Yongjun, Zhang Weihua, et al. DSA Guided Via-Nose Implantation of Intestinal Obstruction Catheter for the Treatment of Advanced Malignant Intestinal Obstruction. [J]. Journal of Nantong University (Medical Science), 2013, 33 (6): 549-550.
- [8] Jin Yinzhi, Wang Qing, Zhang Haiyan. Research Status on the Application of Via-Nose Intestinal Obstruction Catheter [J]. Medical Review, 2012, 18 (21): 3617-3619.
- [9] Jin Yinzhi, Sun Libo, Song Bin. The Therapeutic Effect of Intestinal Obstruction Catheter on Adhesive Intestinal Obstruction [J]. Chinese Journal of Gastrointestinal Surgery, 2009, 12 (6): 580.

Anatomic Variation of Celiac Axis Influencing The D2 Radical Resection For Distal Gastric Cancer

BIN SONG¹, HAI-XIA HU¹, JUAN DU², and YAN-ZHUO SHU¹

¹Department of Gastrointestinal, Colorectal and Anal Surgery, China-Japan Union Hospital of Jilin University, Changchun, Jilin 130033; ²Internal Medicine 2, The Tumor Hospital of Jilin Changchun, Jilin 130012, P.R. China

[Abstract]

Objective: To discuss the anatomic variation of celiac axis influencing the D2 radical operation for distal gastric cancer and its significance.

Method: A retrospective analysis on the anatomic variation of celiac axis influencing operation found in 163 cases of D2 radical resection for distal gastric cancer was made.

Result: 4 cases of anatomic variation of celiac axis influencing operation was found from the 163 cases, 1 case is common hepatic artery originated from left gastric artery, 1 case is left gastric artery was deficiency by 3 branch stomach arteries originating from common hepatic artery, and 2 cases is left gastric artery independently originated from aorta abdominalis.

Conclusion: In the radical resection for gastric cancer, the operator should consider the possibility of coeliac trunk blood vessel variation, carefully anatomize the varying blood vessel to avoid damaging the blood supply for visceral organs.

Keywords: Gastric cancer, operation, celiac axis, anatomic variation

With popularization of standardization of gastric cancer operation, it is necessary to excise the routine lymph nodes in the surrounding blood vessels of celiac axis such as those beside celiac artery and left gastric artery, in front of common hepatic artery, etc. in D2 radical operation for distal gastric cancer. However, the typical type in the celiac axis accounts for 75% [1], and the rest blood vessels all variate in different ways, leading to the risks of damaging the blood vessels and influencing the blood supply in operation. During February 2013 to February 2015, our department totally performed standard D2 radical operation for 163 cases with distal gastric cancer, and found 4 cases with coeliac trunk blood vessel variation influencing operation. Following is a summary of our experience as follows:

1 Clinical data and method

1.1 General materials

The 163 patients with distal gastric cancer who underwent standard D2 radical operation in the surgical department of stomach, intestine, colorectum and anus in China-Japan Friendship Hospital Affiliated to Jilin University during February 2013 to February 2015 were selected as research objects, to observe the variation in the branches originating from the coeliac trunk in D2 radical operation. The included cases all underwent gastroscopy and pathological examination after hospitalized to clarify the diagnosis. Wherein, there were 94 male cases, 69 female cases, whose age was 32-76 and average age was 59.8.

1.2 Method

The D2 lymph node excision was performed during the operation, and the surgeon paid attention to fully exposing celiac axis, common hepatic artery, splenic artery and left gastric artery before ligating the blood vessels. After the duodenum was severed, the surgeon fully exposed gastroduodenal artery and avoid damaging it. At the 3rd, 7th, 14th day after the operation, the liver functions were reexamined, with each index being normal. The liver functions were reexamined at the 3rd, 7th, 14th day after the operation respectively.

2 Result

2.1 Variation type I

1 Case is celiac axis gave out splenic artery and thick left gastric artery, and common hepatic artery originated from 1cm from the starting point of left gastric artery. The treatment method was ligating and severing the common hepatic artery after it was given out by the left gastric artery.

2.2 Variation type II

1 Case is celiac axis gave out splenic artery and common hepatic artery and left gastric artery was missed. Three branch stomach arteries with the similar thickness given out by the common hepatic artery supplied blood to upper part of curvatura ventriculi minor. The treatment method was respectively ligating and severing at the root of 3 branch stomach arteries given out by common hepatic artery.

2.3 Variation type III

2 Case is celiac axis gave out splenic artery and common hepatic artery, and left gastric artery independently originated from abdominal aorta.

3 Discussion

The clinical classification and discussion about variation of celiac axis mostly refer to Michels typing [2]. The Michels type I (normal anatomical type) is the commonest, i.e the celiac axis gives out left gastric artery, splenic artery and common hepatic artery, the common hepatic artery then gives out gastroduodenal artery and arteria hepatica propria, the arteria hepatica propria is divided into left branch of proper hepatic artery and right branch of proper hepatic artery. But there are manifold variations in the blood vessels of coeliac trunk and perigastric blood vessels [3]. The report showed that 10.2% of 1,500 patients undergoing multislice helical CT (MDCT) blood vessel imaging had variation [4], and the practical operation found a higher proportion.

Currently, the standard D2 radical operation for distal gastric cancer, either laparotomy or laparoscopic surgery, mostly adopts the method of going backward to back gap of pancreas on upper edge of pancreas along vagina vasorum of common

hepatic artery to skeletonize left gastric artery and splenic artery, then occluding and dividing left gastric vessels, and excising No.7, 8a, 9, 11p groups of lymph nodes [5]. The key point of this operation is isolation along common hepatic artery and reaching the coeliac trunk, left gastric artery and splenic artery, based on the precondition that coeliac trunk belongs to normal anatomical type. If there are variations in the originating blood vessel branch, such excision mode may damage the blood vessels and unthoroughly excise the lymph nodes.

4 cases of the 163 cases in this group had coeliac trunk variation influencing the operation, accounting for 2.5%. We found that the variations are mainly divided into three types: variation type I (the common hepatic artery originated from left gastric artery). If the left gastric artery is firstly isolated in operation, and left gastric artery is ligated and severed at its root, the common hepatic artery is equal to being ligated and severed synchronously. If the gastroduodenal artery is hurt when duodenum is being isolated, the liver will suffer from interruption of arterial blood supply. Variation type II (left gastric artery is missed), In operation, it is prone to mistake the first branch gastric artery given out by common hepatic artery as left gastric artery trunk, after it is ligated and severed, the surgeon continues to isolate towards coeliac trunk and mistakenly thinks there are no main blood vessels in the gap, so he/she adopts electro-tome or ultrasound knife, which is prone to damage the second and the third branches, thereby leading to massive haemorrhage. Variation type III (the left gastric artery originates from abdominal aorta), in operation, the surgeon should notice that the left gastric artery originates from a deep place, and will fail to find out the root of left gastric artery when reaching coeliac trunk, Now he/she should continue to explore backwards carefully along left gastric artery, avoid damaging coeliac trunk, and after reaching abdominal aorta, he/she should ligate at the root and avoid omitting the lymph nodes at the root of left gastric artery.

As the iconographical technology is advancing, as well as careful operation during operation, preoperative CTA is also a good means to display preoperative variation of blood vessels. CTA can accurately display the invasion of blood vessel by perigastric arteries and their anatomical variation as well as tumor [6], which is favorable for find out anatomical variation before operation to avoid damaging blood vessels and causing interruption of blood supply to visceral organs. Some foreign scholars also apply CTA to preoperative navigation for celioscopic gastric cancer excision [7].

Thus, when popularizing standardization and routinization of gastric cancer operation process, we should not forget variability of blood vessels in coeliac trunk, especially the variation of blood vessels influencing operation. The key to treat the coeliac trunk variation influencing operation is to understand the mutual fusion and migration relation of perigastric aponeuroses, fully expose anatomized plane, finely anatomize blood vessels and fully identify them before ligation. In the meanwhile, application of iconography is also a good means for preoperative judgment.

References

- 1 Saeed M, Murshid KR, Rufai AA, et al. Coexistence of multiple anomalies in the coeliac-mesenteric arterial system[J].Clinical anatomy, 2003,16(1): 30-36 .
- 2 Li Tiegang, Yao Hongliang, Yu Dan, Hepatic Artery Variation in Radical Operation

- for Carcinoma of Stomach and Its Significance [J]. Chinese Journal of Modern Operative Surgery, 2006,10(6):418-419
- 3 Song SY, Chung JW, Yin YH,etal.Celiac axis and common hepatic artery variations in 5002 patients:systematic analysis with spiral CT and DSA[J].Radiology, 2010,255(1):278-288
- 4 Yi Wang,Cheng Cheng,Lu Wang, etal.Anatomical variations in the origins of the celiac axis and the superior mesenteric artery: MDCT angiographic findings and their probable embryological mechanism[J]. European Radiology,2014,24(8):1777-1784
- 5 Xiao Junfeng, Ji Ling, Ruan Xiaojiao, et al. Application of "Pancreas Centre Dissection Method" to Celioscope-assisted Lymph Nodes Excision in Distal D2 Radical Operation for Gastric Cancer [J]. Chinese Journal of Minimally Invasive Surgery, 2013,13(2):119-122
- 6 Zhu Jin, Liu Te, Ling Rennan, et al. Application Value of Displaying Perigastric Arteries and Their Variation by CTA in Preoperative Evaluation on Gastric Cancer. Journal of Clinical Radiology, 2011,30 (8):1150-1152
- 7 Fujiwara M, Kodera Y, Satake H, et al. Navigation for laparoscopic gastrectomy with 3-dimensional computed tomography(3D-CT) [J].Hepato-gastroenterology, 2008,55(85):1201-1205.

Clinical study on the treatment of metastatic malignant bowel obstruction with transgastric Intestinal Obstruction Catheter arrangement small intestinal enterostomy

BIN SONG¹, HAI-XIA HU¹, JUAN DU², and YAN-ZHUO SHU¹

¹ Department of Gastrointestinal, Colorectal and Anal Surgery, China-Japan Union Hospital of Jilin University, Changchun, Jilin 130033;

² Internal Medicine 2, The Tumor Hospital of Jilin Changchun, Jilin 130012, P.R. China

Abstract:

Objective: To explore the value of application of transgastric intestinal obstruction catheter arrangement small intestinal enterostomy in the treatment of metastatic malignant bowel obstruction.

Method: 15 cases of patient with metastatic malignant bowel obstruction who meets the inclusion criteria were performed with a transgastric poke with bowel obstruction catheter for an arrangement of small intestine and intestinal enterostomy; the effect of perioperative period was observed, a post-operation follow-up was conducted to understand the relief of obstruction and survival time.

Result: All 15 patients completed operation, and the operation was successfully done, the operation time was 85–132min (averagely 118.7min); the recovery time of gastrointestinal function was 2–7d, median time was 3.5d; hospitalization time was 8–33d, median hospitalization time was 17.8d; there was one case of death in the peri-operative period. 14 cases were followed up, the follow-up time was 3 - 38 months (averagely 20.7 months), survival time was 3 -38 months, median survival time was 15.4 months, the benefit remission rate was 93.3% (14/15)

Conclusion: A reasonable selection and application of transgastric bowel obstruction catheter arrangement small intestinal enterostomy can relieve the bowel obstruction in part of the patients, and oral feeding can delay the time of reobstruction, extending the survival time.

Keyword: intestinal obstruction catheter; bowel enterostomy; malignant bowel obstruction;

Malignant bowel obstruction refers to the intestinal obstruction caused by primary or metastatic malignant tumor, and it is a common complication and cause of death in patients with advanced tumor [1]. The intestinal obstruction caused by metastatic malignant tumor is often caused by abdominal multiple metastasis, which is easy to relapse after operation, and a conservative treatment has difficulty in relieving the obstruction, which is currently a difficult problem in the treatment of advanced cancer. Based on the role of bowel obstruction catheter in

the conservative treatment and arranged surgical treatment of complex intestinal obstruction [2-3], from June 2007 to August 2014, a transgastric bowel obstruction catheter arrangement small intestinal enterostomy was used to treat metastatic malignant intestinal obstruction in our hospital, and a certain clinical result were achieved , which is now reported as follows:

1 Clinical data

1.1 Case inclusion criteria

Patients with metastatic malignant bowel obstruction were mostly in their advanced stage and had a limited survival time, with a view to the benefits of patients and reduction of complications, the following patients were included into the group: ① devoid of dyscrasia and severe malnutrition, ②, devoid of heart, lung and kidney failure; ③ without a history of surgery in the last 3 months; ④ the patient does not reject bowel enterostomy, and has a strong desire for oral feeding; ⑤ liver, lung, brain and other main functional organs are free of metastasis ; ⑥ the possibility of digestive tract obstruction caused by single tumor is excluded, ⑦ no effect of relief is achieved by trans-nasal indwelling bowel obstruction catheter and other conservative treatments, ⑧ without severe anemia, peritonitis, fever and serious infection.

1.2 General data

All 15 cases were diagnosed before operation as conforming to the inclusion criteria. Among them there were 7 cases of male and 8 cases of female; with an age of onset ranging from 45 to 67, median age was 56.7 years old. The interval between the development of symptom of bowel obstruction and the operation was 4 days to 1 month. 9 Cases (4 cases of patient had been performed with ovarian cancer radical resection, 3 cases had been performed with radical resection of gastric cancer, 2 cases had been performed with radical resection of colorectal cancer) had been performed with laparotomy to conduct a resection of primary tumor and a post-operative chemotherapy, 1 case of patient was in a state of post-operation of local resection of skin cancer , 5 cases were found to be primary attack case, (3 case of extensive metastasis of ovarian cancer in abdominal cavity, 2 case of extensive metastasis of colon cancer in abdominal cavity). Before the operation, 5 cases were found to have a moderate anemia , and 4 cases were found to have a mild anemia. Preoperative albumin : 28 ~ 36g/L. Routine examinations like abdominal ultrasonography, whole-abdomen CT, gastroscopy and colonoscopy were performed, the female patients underwent a gynecological ultrasound examination, and a small intestine radiography was performed through trans-nasal indwelling bowel obstruction catheter. The intestinal obstruction catheter is made of three cavities (front airbag, Posterior airbag) and the intestinal obstruction catheter being used was the 3-cavity-2-bag (front air bag, back air bag) bowel obstruction catheter that is produced by Japanese CREATEMEDIC company.

1.3 Method

1.3.1 Preoperative basic treatment

The patients were administered with fluid infusion, correction of anemia, correction of hypoproteinemia , nutritional support, enema, correcting acid-base imbalance and ion disorder before the operation, all cases before operation were corrected to the point where their hemoglobin was more than 90g/l and albumin more than 35g/l. After admission they were immediately administered with a small intestine decompression after the bowel obstruction catheter was indwelled, the bowel obstruction catheter passed through smoothly in 10 cases, to the depth of 2.3 ~ 2.9 m; and the intestinal obstruction catheter reached to the depth of 60cm and 65cm in two cases respectively. In 15 cases of patient, the abdominal distension and

abdominal pain were relieved after decompression, and the duration of conservative treatment was 5~10d, a meglumine diatrizoate radiography was carried out through the bowel obstruction catheter and it was found that, the distal end was still obstructed. On the day of operation, the intestinal obstruction catheter was withdrawn, and disinfected by povidone iodine for standby use.

1.3.2 Intraoperative operation

All cases were performed with a general anesthesia, all of them were poked through the greater curvature of stomach in order to arrange the intestinal obstruction catheter to the position of small intestinal enterostomy, it is required that, more than 2m of small intestine that is free of obvious stenosis and able to allow the 10mL water bag of bowel obstruction catheter to pass through be retained. ① treatment in the case of absence of pylorus and duodenal stenosis. A lateral incision with a diameter of about 0.5cm was performed through the greater curvature of stomach, and the head of intestinal obstruction catheter was inserted into the gastric cavity through the abdominal wall. The intestinal obstruction catheter was introduced into the small intestine through the pylorus and duodenum till it reached the fully obstructed part, and the intestinal juice was sucked by negative pressure suction at the same time. If the length of intestine at the proximal end of obstruction is more than 2m, then the normal small intestine 10cm above the obstruction is raised through the lateral abdominal wall out of the stoma, and at the same time the intestinal obstruction catheter was retained in intestine cavity.

If the length of intestine at the proximal end of obstruction is less than 2m, the intestinal canal at the fully obstructed position should be resected and an anastomosis should follow, or the obstructed position should be put aside to perform a short-circuit operation, followed by a small intestine enterostomy. Finally, the gastric incision should be embedded with double-pouch method and fixed in the abdominal cavity and at the abdominal wall skin. ② the treatment in the case of presence of pylorus and duodenal stenosis. A gastrointestinal anastomosis should be firstly performed, then the bowel obstruction catheter should be introduced into the small intestine through gastrointestinal anastomotic stoma till it reach the fully obstructed part, the remaining treatment is the same as the treatment in the case of absence of pyloric and duodenal stenosis.

1.3.3 Result judgment criteria

The criteria of benefit and remission should be subject to the patient being capable of normal oral feeding, free of relapse of bowel obstruction and surviving for more than 60 days.

1.4 Result

1.4.1 Intraoperative condition

Operation duration: 85~132min (averagely 118.7min). 4 cases of patient were performed with small intestine enterostomy alone, 4 cases of patient were performed with resection of small intestine and anastomotic small intestine enterostomy; 3 cases were performed with small intestine putting-aside and side-to-side anastomotic small intestine enterostomy; 4 cases of patient with ovarian cancer complicated with duodenal stenosis and jejunum proximal stenosis were performed with gastrointestinal anastomosis small intestine enterostomy.

1.4.2 Postoperative recovery

The recovery time of gastrointestinal function for this group of patients was 2~7d, median time was 3.5d; hospitalization time was 8~33d, median hospitalization time was 17.8d.

1.4.3 Death and complication

One case of patient died from a postoperative small intestine fistula and disruption of wound 21 days after operation; 5 cases were found to have incision infection; 1 case was found to have infection at the position of stoma, and exudation of gastric juice was found, which was relieved after repeated dressing change; 1 case was found to have intra-abdominal abscess, which was relieved by drainage and anti-infection.

1.4.4 Postoperative follow-up result

The patients were followed up by outpatient service and telephone follow-up on month 1, 2, 3, 6, 12, 24 and 36 after operation respectively, with an exception that 1 case of patient died on day 21 after operation, the remaining 14 cases of patient were all followed up, follow-up duration was 3 - 38 months (averagely 20.7 months). Survival time was 3 -38 months, median survival time was 15.4 months. 2 death was reported during the follow-up period, as a result of re-obstruction. In this group, 14 cases of patient were benefited and remitted and the benefit remission rate was 93.3%.

2 Discussion

Malignant intestinal obstruction is a common complication in the late stage of malignant tumor, including the intestinal obstruction caused by primary or metastatic malignant tumor. Different from the primary malignant intestinal obstruction, the patients with metastatic malignant bowel obstruction usually have multiple lesions, which are often scattered over the whole abdominal cavity, the obstruction mostly occurs at the small intestine, it is common in the terminal stage of ovarian cancer, colorectal cancer, gastric cancer and other tumors, and is currently a difficult problem faced in the treatment of advanced cancer [6].

The current treatment methods mainly include traditional surgery, placement of extendable stent, drug therapy, percutaneous endoscopic gastrostomy, nasogastric tube drainage and total parenteral nutrition, the traditional operations usually prefer stage-1 intestinal anastomosis and bypass surgery, according to the conservative view, the operation is only suitable for mechanical obstruction and/or the obstruction with a limited tumor and single site, and for the patients who may benefit from further chemotherapy and anti-tumor therapy; extendable Stent therapy is usually used in the treatment of the pyloric and proximal small intestine and colonic obstruction, but it is regarded as a stent therapy contraindication for the patients with multiple obstructions and extensive metastasis in abdominal cavity [6]. Long-term intravenous nutrition, which is expensive and at a high risk of infection, blood glucose disorders and other complications, can only ensure the supply of nutrition and can not solve the obstruction problems. The application of Anti-secretion drugs, antiemetics, glucocorticoid and other drugs can only mildly alleviate abdominal distension, reduce some of the digestive fluids, and must be used in combination with gastrointestinal decompression and intravenous nutrition.

This study used intestinal obstruction catheter to perform a nasal small intestine decompression before the operation in order to relieve the patient's abdominal distension, and to avoid emergency operation, preoperative adjustment time was 5~9d, which has won time for correcting the hypoproteinemia. In this group, the level of preoperative albumin was maintained above 35g/l, and at the same time a reasonable supplement of nutrition was administered to correct anemia and electrolyte disorder, helping to improve the overall physiological function disorder caused by obstruction. Some studies have shown that an albumin level that is lower than 35g/l can be used as a poor prognostic factor for patients with malignant intestinal obstruction, suggesting that patients with malignant intestinal obstruction whose albumin level is

less than 35g/L should be given a targeted treatment as early as possible, which may improve the overall prognosis of patients.

12 cases did not develop intestinal obstruction again till their death, suggesting that, the retention of bowel obstruction catheter plays a role of stent in the intestinal cavity , to a certain extent, it has slowed the growth of tumor towards the cavity , kept the proximal intestinal tract basically unobstructed and postponed the re-stenosis . Literature has reported that the throat discomfort, pain and other complications brought by the nasal indwelling catheter are 98.1% likely to happen, some patients have a serious reaction and can not tolerate it. We use trans-gastric enterostomy indwelling catheter to avoid such complications that affect the quality of life of patients, and it can be carried for a long time and doesn't fall off.

In a word, metastatic malignant bowel obstruction is a clinical difficulty, it is necessary to combine the patients ' actual conditions and strictly grasp the indications; rationally choosing the use of transgastric bowel obstruction catheter arrangement small intestinal enterostomy can relieve part of the patients from obstruction and promote oral feeding while extending the time of re-obstruction and survival time and improving the survival quality.

References

1. Selby D, Wright F, Stilos K, et al. Room for improvement? A quality-of-life assessment in patients with malignant bowel obstruction. *Palliat Med*, 2010, 24(1): 38-45.
- 2 Niu lujie, Sun libo, Zhang Bing et al . Clinical analysis of 38 cases where the bowel obstruction catheter is used to treat low bowel obstruction . *Chinese Journal of General Surgery*, 2010, 25(1): 79-80.
- 3 Jin Yinzhi, Sun Libo, Song Bing . the therapeutic effect of bowel obstruction catheter on adhesive intestinal obstruction . *Chinese Journal of Gastrointestinal Surgery*, 2009, 12(6): 580.
- 4 Dolan EA. Malignant bowel obstruction: a review of current treatment strategies. *Am J Hosp Palliat Care*, 2011, 28(8): 576-582.
- 5 Yu Shiyong, Wang Jiejun, Wang Jinwan et al . expert consensus on the treatment of patients with advanced cancer complicated with bowel obstruction . *Chinese Journal of Tumor*, 2007, 29(8): 637-640.
- 6 Zhang Qifan, Zheng Hongqun, Sun Lingyu . Diagnosis and treatment of malignant intestinal obstruction caused by peritoneal metastasis of gastric cancer and nutritional support strategies. *Chinese Oncology Clinical*, 2014, 41(12): 749-752.
- 7 Hisanaga T, Shinjo T, Morita T, et al. Multicenter prospective study on efficacy and safety of octreotide for inoperable malignant bowel obstruction. *Jpn J Clin Oncol*, 2010, 40(8): 739-745.
- 8 Wang Aman, Ning Zheng, Zhou Tao et al . Prognostic analysis of advanced tumors complicated with malignant intestinal obstruction. *Modern Oncology medicine*, 2014, 22(2): 379-383.
- 9 Barbot AC, Mussault P, Ingrand P, et al. Assessing 2-month clinical prognosis in hospitalized patients with advanced solid tumors. *J Clin Oncol*, 2008, 26(15): 2538-2543.
- 10 He Xuehong, Mao Aiwu, Fang Shiming et al . Treatment of benign and malignant intestinal obstruction by DSA-guided nasal inserted intestinal obstruction catheter. *Western medicine*, 2013, 25(10): 1548-1550.

Equilibrium and Thermodynamic Studies of Isoniazid Adsorption by Montmorillonite

E. Carazo^{1*}, A. Borrego-Sánchez^{1,2}, F. García-Villén¹, R. Sánchez-Espejo¹,
P. Cerezo¹, C. Aguzzi¹, C. Viseras^{1,2}

¹ Department of Pharmacy and Pharmaceutical Technology, School of Pharmacy, University of Granada, Campus of Cartuja s/n, 18071, Granada, Spain ecarazogil@ugr.es

² Andalusian Institute of Earth Sciences, CSIC-University of Granada, Avda. de Las Palmeras 4, 18100, Armilla (Granada), Spain

Keywords: Isoniazid; Montmorillonite; Adsorption; Thermodynamics; Nanohybrids

1 Abstract

The properties of montmorillonite (MMT) to act as a natural nanocarrier of drugs has been fully reported. Isoniazid (INH), a first-line tuberculostatic drug, was intercalated into this inorganic matrix and the equilibrium and thermodynamic aspects of the adsorption process were evaluated. Adsorption experiments were conducted at different times and temperatures in aqueous medium. The overall adsorption process was the result of two simple processes: drug adsorption on the activated sites of montmorillonite followed by a slight precipitation phase of drug molecules over the adsorbed monolayer. Formation of the nanohybrid was spontaneous, exothermic and exoentropic, obtaining an increase in the thermodynamic stability of the system.

2 Introduction

The use of nanotechnology [1] and in particular of clay minerals [2] aimed to alleviate the difficulties derived from the tuberculosis (TB) treatment has evoked great interest worldwide in the current years. In particular, one of the first-line tuberculostatic drugs, isoniazid (INH) has been effectively loaded in tubular clay minerals like halloysite [3] and fibrous clay minerals like palygorskite [4]. Montmorillonite (MMT) is a layered silicate of the smectite group with alumina octahedral sheets between silica tetrahedral sheets with countless applications in the biomedical field like drug delivery and wound healing [5]. With these premises, aim of this work was to investigate equilibrium and thermodynamic aspects of the adsorption of isoniazid onto a pharmaceutical-grade MMT as a required research preceding to the elaboration of prolonged release systems of the drug.

3 Materials and Methods

3.1 Materials

Isoniazid (INH) was purchased from Sigma Aldrich (Spain); Pharmaceutical grade montmorillonite (Veegum HS[®]) (MMT) was kindly gifted by Vanderbilt Company (USA).

3.2 Adsorption Studies

Adsorption experiments were performed following [3] to obtain equilibrium isotherms at different times and temperatures. Briefly, known amounts of MMT were dispersed into INH aqueous solutions with initial concentration (C_0) ranging from 0,05 to 0,5 mol / L , for different times (24 h, 48 h and one week) and temperatures (20, 30, 40 \pm 0.1°C). The resulting dispersions were centrifuged and the equilibrium concentration (C_e) of the drug in the supernatant was determined by UV spectroscopy (UV-Vis spectrophotometer Lambda 25, Perkin Elmer, S) at 262 nm. The difference between C_0 and C_e was used to calculate the amount of drug retained per gram of clay (expressed as n^s : mol of INH / g of MMT). Non-linear fitting of the data was performed [3] using the software packaging TableCurve 2D[®] (Systat Software Inc., UK) and kinetic and thermodynamic parameters were determined.

4 Results and Discussion

Equilibrium isotherms obtained at different times and temperatures (data not shown) revealed that contact time did not significantly influence the adsorption results, allowing considering 24 h time as time sufficient enough to reach equilibrium. The isotherms fit the following equation, previously proposed to describe the adsorption of drugs to inorganic solid sorbents [3,4] (Equation 1):

$$n^s = n_1^s + n_2^s = \sum_{i=1}^{i=2} \frac{K_i \cdot n_m^s \cdot C}{K_i \cdot C + 1} + k_f C^m \quad (1)$$

n^s = moles of drug adsorbed per gram of sorbent, n_1^s = adsorption at the active surface sites of the sorbent, n_2^s = precipitation of drug molecules over the adsorbed monolayer, n_m^s = monolayer retention capacity, C = equilibrium concentration, k_i = kinetic equilibrium constant, k_f = kinetic precipitation constant, n = partial order of the process respect to the concentration (C) and m = constant.

The global adsorption process is composed by the sum of two simple processes (adsorption of INH on the activated sites of MMT (n_1^s) followed by a slight precipitation of INH over the adsorbed monolayer (n_2^s)). Average experimental data and theoretical calculated curves (adsorption, precipitation and the sum of both processes) are shown in Figure 1.

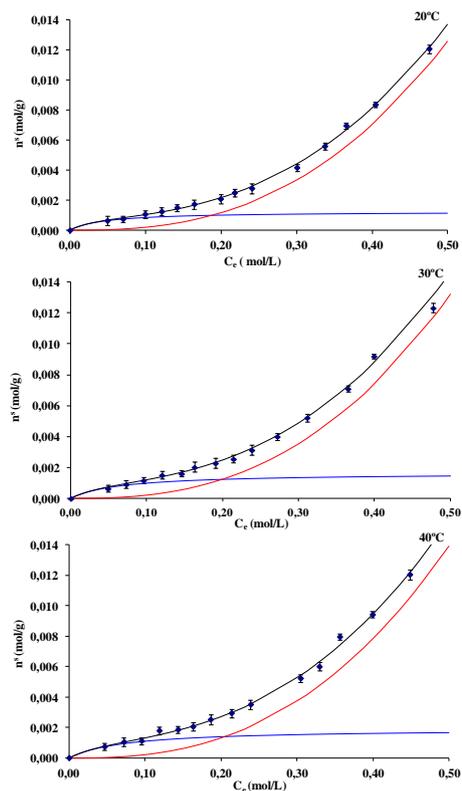


Figure 1. Equilibrium isotherms of INH by MMT at different temperatures (mean values \pm s.d.; $n = 3$).

The calculated fitting parameters (Table 1) described adequately the experimental results with correlation coefficient ≥ 0.99 . Monolayer retention capacity (n_m^s) slightly increased with temperature, whereas kinetic equilibrium constant (k_i) significantly decreased. The increase in temperature makes easier the dehydration of the drug molecules previous to their adsorption but concomitantly increase their mobility. K_f values are very low, suggesting a slight precipitation over the adsorbed drug monolayer.

Table 1. Fitting parameters of Equation 1 for the retention of INH by MMT at different temperatures and corresponding correlation coefficients (mean values \pm s.d.; n = 3).

T° (C)	n^s_m (mmol / g)	k_i	k_f	m	r^2
20	1,143 \pm 0,0011	2,131 \pm 0,3431	0,032 \pm 0,0112	2,603 \pm 0,1234	0,998
30	1,502 \pm 0,0012	1,502 \pm 0,4122	0,033 \pm 0,0234	2,587 \pm 0,2454	0,994
40	1,764 \pm 0,0015	1,368 \pm 0,4751	0,048 \pm 0,0346	2,568 \pm 0,2987	0,997

Thermodynamic parameters including differential adsorption enthalpy ΔH° (kJ / mol) differential adsorption entropy ΔS° (kJ / mol K) and differential activation energy ΔG° (kJ / mol), were evaluated to confirm the nature of the adsorption of INH by MMT. Firstly, ΔH° and ΔS° were calculated via the linearization of the Eyring equation (Eq. 2) by plotting $\ln(k/T)$ vs $(1/T)$ using the value of k_i previously determined in Eq. 1:

$$\ln \frac{k_i}{T} = \ln \frac{R}{Nh} + \frac{\Delta S^\circ}{R} - \frac{\Delta H^\circ}{R} \times \frac{1}{T} \quad (2)$$

(k_i = kinetic equilibrium constant, R = ideal gas constant, N = Avogadro's number, h = Planck's constant, T = temperature (K)).

The calculated values are given in Table 2, which also includes the ΔG° (kJ/mol) values calculated via the expression: $\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ$.

Table 2. Thermodynamic functions for the MMT/INH interaction

T (K)	ΔH° (kJ / mol)	ΔS° (kJ / mol K)	ΔG° (kJ / mol)
293	-17,291 \pm 1,7780	-0,061 \pm 0,0011	-2,731 \pm 0,0014
303			-2,521 \pm 0,0012
313			-2,505 \pm 0,0021

Adsorption of INH over MMT was an exothermic process. One possible interpretation of the exothermicity of the process was that INH molecules were solvated in water and to become adsorbed they required at least a partial decomposition of their hydration shell, a process that required energy. However, the exothermicity associated with the adsorption of INH molecules on MMT was higher than the dehydration energy so the overall energy balance was found to be exothermic. ΔG° negative values stated that the process was spontaneous under the conditions applied and the reversible adsorption-desorption process moved towards the net adsorption. Besides, the

negative value of ΔS° corresponded to the decrease of randomness of the INH adsorbed molecules with respect to those in dissolution.

5 Conclusions

Equilibrium isotherms fit to two simple processes: adsorption of INH on the activated sites of MMT followed by a slight precipitation of the drug on the clay mineral surface. Calculation of the fitting parameters of the equation that defines the processes and the thermodynamic functions reveals that: the INH retention capacity by the nanoclay increases with increasing temperature; the precipitation step is of low intensity and the formation of the INH/MMT nanohybrid fits to an exothermic, exoentropic and thermodynamically spontaneous process.

Acknowledgements

This study was supported by the Projects CGL2016-80833-R and RNM-1897, group CTS-946 and Predoctoral Grant FPU13/04765.

References

1. Bullo Saifullah, M.Z.B.H., Al Ali, S.H.H.: Controlled-release approaches towards the chemotherapy of tuberculosis. *Int. J. Nanomed.*, **7** (2012), 5451. doi:10.2147/IJN.S34996
2. Carazo, E., Borrego-Sánchez, A., Aguzzi, C., Cerezo, P., Viseras, C.: Use of clays as nanocarriers of tuberculostatic drugs. *Curr. Drug Deliv.*, **14**(7) (2017), 902-903. doi:10.2174/1567201813666160714160727
3. Carazo, E., Borrego-Sánchez, A., García-Villén, F., Sánchez-Espejo, R., Aguzzi, C., Viseras, C., Sainz-Diaz, C.I., Cerezo, P.: Assessment of halloysite nanotubes as vehicles of isoniazid. *Colloids Surf. B Biointerfaces*, **160** (2017), 337-344. <https://doi.org/10.1016/j.colsurfb.2017.09.036>.
4. Carazo, E., Borrego-Sánchez, A., García-Villén, F., Sánchez-Espejo, R., Viseras, C., Cerezo, P., Aguzzi, C.: Adsorption and Characterization of palygorskite-isoniazid nanohybrids. *Appl. Clay. Sci.*, (2017), <https://doi.org/10.1016/j.clay.2017.12.027>
5. Aguzzi, C., Sandri, G., Bonferoni, C., Cerezo, P., Rossi, S., Ferrari, F., Viseras, C.: Solid state characterisation of silver sulfadiazine loaded on montmorillonite/chitosan nanocomposite for wound healing. *Colloids Surf. B Biointerfaces*, **113** (2014) 152-157. <https://doi.org/10.1016/j.colsurfb.2013.08.043>.

The PERSON project: A brain computer interface serious game for neurorehabilitation

Alfonso Monaco¹, Nicola Amoroso^{1,2}, Roberto Bellotti^{1,2}¶, Paolo Da Pelo²,
Domenico Diacono¹, Gianluca Sforza¹, Sabina Tangaro¹¶

¹ INFN, Sezione di Bari, Via A. Orabona 4, 70125 Bari, Italy

² Department of Physics 'Michelangelo Merlin', University of Bari 'Aldo Moro', Via
G. Amendola 173, 70126 Bari, Italy

¶These authors contributed equally to this work.

Abstract. Neurodegenerative diseases represent an important reason of disability and premature death among older people worldwide [1]. According to the World Alzheimer Report 2016, 47 million people are affected by dementia and it is expected that over 131 million people will develop dementia by 2050. Therefore these diseases constitute an important public health problem with considerable economic and social health impact. In 2008 in the Europe countries costs for these disease amounted to 160 billion with an expected 43% increase by 2030. Accordingly, the cost of care for patients with a form of dementia is expected to increase dramatically. It is well known how a brain disease could have a devastating effect not only on the people directly affected, but also on their families and yields huge costs, often not sustainable, for health-care systems. Early and accurate diagnosis has great potential to reduce the costs related to care and living arrangements as it gives patients access to supportive therapies that can help them maintain their independence for longer and delay institutionalization. To date there is no cure for neurodegenerative diseases, but treatment can still help in reducing symptoms and providing a better quality of life. Therefore it is necessary to define tools able to treat subjects suffering from the early symptoms, in order to give priority to a home environment. In this way the curative treatment does not burden on the hospital and the patient's psychological disorientation. Recent studies have shown the efficacy of using video games to improve cognitive processes caused by aging physiological effect and neurodegenerative disease avoiding drug treatment [2, 3]. In this context the PERSON (PERvasive game for perSONalized treatment of cognitive and functional deficits associated with chronic and Neurodegenerative diseases) project [4], funded by the Apulia Region on behalf of public and private institutions, developed a rehabilitative tool. This platform includes a serious game connected to brain computer interface systems composed by an EEG interface (electroencephalography paradigm through Brain Computer Interface) and haptic devices. The project concerns a strategic synergy between public health, research and technology to develop a mobile device, that involves mild cognitive impairment patients. The device is based on Virtual Reality and software solutions for personalized medicine basing on an open source cloud infrastructure. In particular PERSON capabilities are:

- A serious game based on Virtual Reality that interfaces with patient;
- the definition of algorithms and models which extract from data personalized medical responses, such as diagnosis and patient assessment;
- the customization of these tools in terms of software as a service solutions (SaaS).

In patients with brain impairment or with dementia, the PERSON serious game, through exercises at various levels of complexity, can stimulate concentration, imagination, with diverse problem-solving strategies designed to perform the required task. Exercises are designed to allow the patient to follow the movement of objects across the virtual room, with increasingly complex levels of difficulty, by introducing and monitoring, for example, elements of disturbance or by changing the speed of execution, in order to progressively increase the emotional and cognitive pressure. This boosts the patient motor and cognitive reaction that are recorded, through EEG interface and haptic devices, and then analyzed. The PERSON serious game was designed to improve neurocognitive abilities through simple task as simple math tasks, simple text games, image recognition, Stroop test, Posner Task. The serious game interface is connected via a 32 channels wireless EEG with active electrodes. The main EEG component sought is the Event Related Potentials (ERP) P300 [5], a signal that indicates the oriented attention and working memory toward target stimuli. These stimuli are usually presented by means the so called Oddball paradigm [6]. In this paradigm, a target stimulus is rarely presented among several frequent stimuli in a random way. This generates a positive wave obtained through an averaging, presenting approximately 300 ms after the target stimulus [7].

Furthermore we designed and implemented data mining and predictive models to evaluate the clinical status of subjects. These tools used genetics algorithms to analyze cognitive EEG responses and specifically ERP and Evoked Potentials [8]. In particular we have developed a global approach for the inference of single trial latencies of Event-Related Potentials to find the global maximum. Furthermore our method provides for each patient a personalized score assessing its cognitive impairment: this information can be crucial if combined with other meta data to understand the clinical history and, hopefully, the disease outcomes. We offer these workflows as a service according to the cloud paradigm exploiting the container (Docker) technology [9]. In particular we developed a software component to connect the serious game to a cloud infrastructure (this software component is called cloud-backend). The cloud backend system takes care of the submission in a distributed cloud infrastructure of all of jobs required by the serious game interface, their monitoring and bookkeeping. The system capability hides the complexity of operating in a heterogeneous and distributed computational environment. The aims of cloud-backend are: 1. to perform main tasks to the advancement of the game; 2. to store data and informations produced by the game applications. The system consists of four open-source software elements: Mesos, Chronos, Docker and Swift. Chronos [10] contains the list of tasks to be executed inserted by the serious game system through a WEB services interfaces (REST based). The middleware Mesos [11] submits tasks, when

there are task to be executed in Chronos. Furthermore Mesos takes care of the job scheduling and game system notification. Tasks are executed by means of a Docker container and final results are stored in the object storage Swift [12]. Within PERSON project we have developed an open source and innovative platform that represents a method to improve the quality of life of patients, with mild cognitive impairment, by using a complex and not invasive technology.

References

1. Brettschneider, J., Del Tredici, K., Lee, V.M., Trojanowski, J.Q.: Spreading of pathology in neurodegenerative diseases: a focus on human studies. *Nat Rev Neurosci.* 16(2), 109–20 (2016).
2. McCallum, S., Gamification and serious games for personalized health. *Stud Health Technol Inform.* 177, 85–96 (2012).
3. Bowes, J., Brown, A.J., Hamon, J., Jarolimek, W., Sridhar, A., Waldron, G., Whitebread, S. Reducing safety-related drug attrition: the use of in vitro pharmacological profiling. *Nat Rev Drug Discov.* 11(12):909-22, (2012).
4. <http://progettoperperson.it/>
5. Luck, S.J. An Introduction to Event-Related Potentials and Their Neural Origins. *An Introd. to event-related potential Tech.* 250, (2005).
6. Polich, J. and Margala, C. P300 and probability: Comparison of oddball and single-stimulus paradigms. *Int. J. Psychophysiol.* 25 (2):169176, (1997).
7. Luck, S.J. An Introduction to the Event-Related Potential Technique. *Monogr. Soc. Res. Child Dev.* 78(3):388, (2005).
8. Da Pelo, P., De Tommaso, M., Monaco, A., Stramaglia, S., Bellotti, R., Tangaro, S. Trial latencies estimation of event-related potentials in EEG by means of genetic algorithms. *Journal of Neural Engineering.* 15(2), (2018)
9. <https://www.docker.com/>
10. <https://mesos.github.io/chronos/>
11. <http://mesos.apache.org/>
12. <https://www.swiftstack.com/product/openstack-swift>

Cytotoxicity of Secnidazole cocrystals on cells infected with *Trypanosoma cruzi*

Barrientos Salcedo Carolina ^a, Vichi Ramírez Micheel Merari ^a, Edgar López López ^a, Jorge G. Domínguez Chávez^a, Karina Mondragón Vázquez^a, Misaela Francisco-Márquez^b, C. Ignacio, Sainz-Díaz^c, Bertha J. Espinoza Gutiérrez^d, Ignacio Martínez-Martínez ^dand Soriano Correa Catalina^e.

^a Laboratorio de Química Médica y Quimiogenómica, Facultad de Bioanálisis-Veracruz, Universidad Veracruzana, C.P. 91700, Veracruz, Mexico

^b Instituto Politécnico Nacional, UPIICSA, Av. Té #950, Col. Granjas México, C.P. 08400 Mexico City, Mexico

^c Instituto Andaluz de Ciencias de la Tierra, CSIC-Universidad de Granada, Av. de las Palmeras, 4, 18100-Armilla, Granada, Spain.

^d Laboratorio de Tripanosomiasis americana, Departamento de Inmunología, IIB-Universidad Nacional Autónoma de México, Edificio B 1er Piso / Oficina B-113 , Laboratorio B-114.

^e Química Computacional, Facultad de Estudios Superiores Zaragoza, Universidad Nacional Autónoma de México (UNAM), Iztapalapa, C.P. 09230 Mexico City, Mexico

The *Trypanosoma cruzi* parasite is the causal agent of Chagas disease; with several million infected in the world; and for whose treatment there are only two drugs, and whose secondary reactions and genotoxicity are very high; In addition, it is important to highlight that the main pharmaceutical companies have no interest in synthesizing and commercializing trypanocidal drugs. The reuse of drugs already known for other uses, has contributed to improve the current therapeutic; and since secnidazole has a known antimicrobial and antiparasitic effect against amoebae, but is not used against *T. cruzi*, but in both parasites there is a NADPH-mediated nitroimidazole reduction system, in this thesis it was hypothesized that the co-crystal secnidazole has low cytotoxicity and trypanocidal activity in infected cells; which will allow to propose it as an alternative for the treatment of the chronic stage of the disease. The similarity between the coupling sites and the characteristics of the interactant amino acids between Beznidazol, Nifurtimox and Secnizadol with Cruzipain and Nitroreductasa proteins, made by us with docking, open the possibility to use Secnidazole as a new therapeutic agent in the treatment of Chagas disease (figure 1). In addition to the above, cultures of VERO cells (ATCC CCL-81), cultures of *T. cruzi* epimastigotes, and cellular infection by this parasite were carried out. The previously known drugs, benznidazole (BNZ) and nifurtimox (NFX), as well as secnidazole (SEC) and secnidazole in co-crystal (CO-SEC) were used as controls. The inhibitory doses 50 (IC50) were determined, being 151.2 for NFX, 53 for BNZ, 150.3 for SEC and 251.7 for COSEC. The percentage of cytotoxicity was of BNZ 59.67%, NFX

36.14%, SEC 24.37% and COSEC 28.57%. And it was found that the cocrystal was active against infected cells. Therefore, it is concluded that the secnidazole cocrystals are not cytotoxic, for the model of cells infected with *T. cruzi*, but it is effective against the parasite (Figure 2).

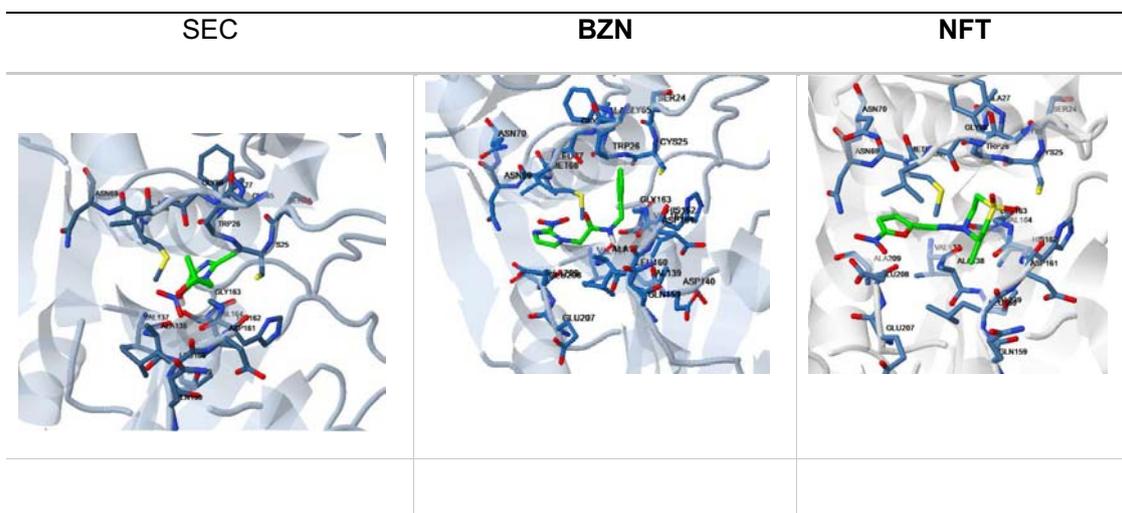


Figure 1. Molecular Docking (Nitroimidazoles vs Cruzipaina).

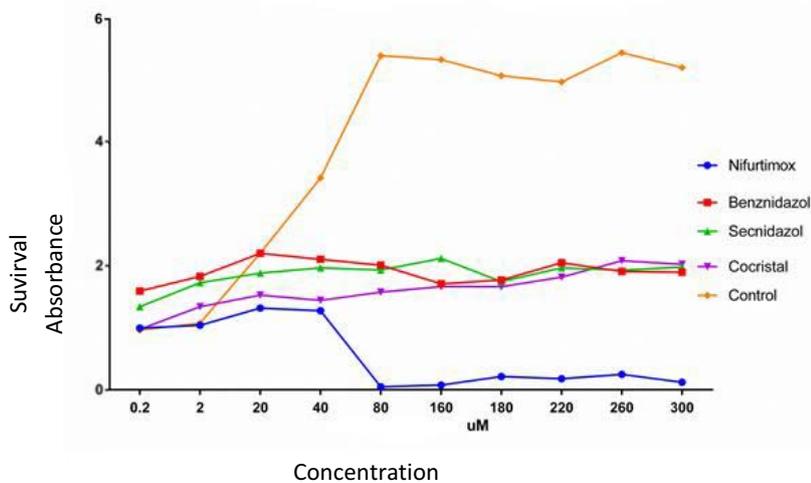


Figure 2. Cytotoxicity VERO ATCC (CCL-81) cells infected with *Trypanosoma cruzi* (0.2 μ M a 300 μ M)

Experimental Design for Optimizing the Properties Antimicrobial Chitosan Derivatives and Investigation of the Structure Activity Relationship

P. Sahariah¹, B. S. Snorradóttir¹, Martha Á. Hjálmarsdóttir², Ólafur E. Sigurjónsson^{3,4},
*M. Másson¹.

¹ Faculty of Pharm. Sci.; University of Iceland, Hofsvallagata 53, IS-107 Reykjavik, Iceland

²Dep. of Biomedical Science, Fac. of Medicine, University of Iceland, Stapi, Reykjavík, Iceland

³The REMoDEL Lab, The Blood Bank, Landspítali Univ. Hosp., Snorrabraut 60, 105 and the Inst. of Biomed. and Neural Eng., Reykjavik University, Menntavegur 1, Reykjavik, Iceland

mlasson@hi.is

Abstract. Design of Experiment (DOE) was used to guide the synthesis of 14 chitosan derivatives based on varying degree of substitution for three substituents, *N,N,N*-trimethyl, *N*-acetyl and *N*-stearoyl. These were assayed for antimicrobial activity, toxicity and solubility. Fitting to response surface model showed that materials with DS for trimethylation in the range 0.45-0.65, acetylation in the range 0.08-0.33 and steroylation in the range 0.22-0.29 were capable of showing high antimicrobial activity, high solubility and low hemolytic activity

Keywords: DOE, -Chitosan, Antimicrobial, Response Surface Modelling

Chitosan is biopolymer derived from marine sources. It has some unique biological properties and has been used in medicine to stimulate tissue regeneration, for gene delivery and as antimicrobial agent[1]. It has also commonly used in for preparing biocompatible nanoparticles and as a starting material for the synthesis of derivatives and nanoconjugates [2] intended for antimicrobial, regenerative, diagnostic and drug delivery applications.

Our research focused chitosan conjugates antibacterial and as delivery systems and agents. For this purpose, we have developed tertbutyl dimethyl silyl (TBDMS) protection strategy to allow selective conjugation of various active moieties to the 2-amino-groups in the polymer backbone[3]. With this method we can also have full control of the degree of substitution. Furthermore, we have used this synthesis approach for synthesis of highly active antimicrobial chitosan derivatives and conjugates. Good control of the synthesis has enabled detailed studies to show that activity is much more sensitive small changes in the chemical structure than previously thought.

In the current work we have, for the first time, used a statistical experimental design, or Design of Experiment (DOE) to plan a study of the structure activity relationship (SAR) antimicrobial chitosan conjugates and to optimize the structures with regard to antimicrobial activity as well as low toxicity and good solubility.[4]

Specialized software (MODDE) with D-optimal design capabilities was used to create a library of chitosan derivatives with optimal structural variation in order to conduct a detailed investigation of the structure-activity relationship. The derivatives contain three substituents: *N,N,N*-trimethylamine, *N*-acetyl and *N*-stearoyl at different degrees of substitution (DS) on the 2-amino group of chitosan. The design matrix consisted of 14 target materials that were synthesized in 'one-pot synthesis' using TBDMS-chitosan as the precursor to allow precise control of the DS (Fig 1).

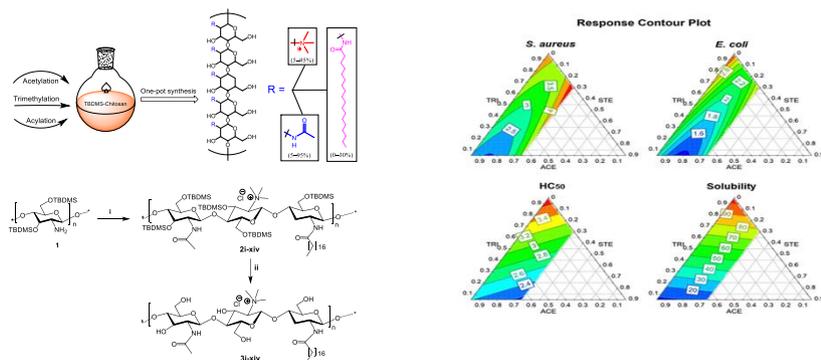


Fig.1. (Right) Scheme Synthetic route for 'one-pot synthesis' of chitosan derivatives. (Left) Response contour plot for the antibacterial activity (*S. aureus* and *E. coli*), hemolytic activity and solubility of the chitosan derivatives

The antibacterial activity (MIC) towards the Gram positive bacteria *Staphylococcus aureus* and the Gram negative bacteria *Escherichia coli*, hemolytic activity (HC₅₀) towards human red blood cells and solubility of the chitosan derivatives were used as the responses in the model. The response surface model was refined by removing the interaction terms to improve the statistical significance and predictive power of the model. The investigation showed that materials with DS for trimethylation in the range 0.45-0.65, acetylation in the range 0.08-0.33 and stearoylation in the range 0.22-0.29 were capable of showing high antimicrobial activity, high solubility and low hemolytic activity

References

- [1] P. Sahariah, M. Masson, Antimicrobial Chitosan and Chitosan Derivatives: A Review of the Structure-Activity Relationship, *Biomacromolecules* 18(11) (2017) 3846-3868.
- [2] P. Sahariah, K.K. Sorensen, M.A. Hjalmarsdottir, O.E. Sigurjonsson, K.J. Jensen, M. Masson, M.B. Thygesen, Antimicrobial peptide shows enhanced activity and reduced toxicity upon grafting to chitosan polymers, *Chem. Commun.* 51(58) (2015) 11611-11614.
- [3] Ö.V. Rúnarsson, C. Malainer, J. Holappa, S.T. Sigurdsson, M. Másson, tert-Butyldimethylsilyl O-protected chitosan and chitooligosaccharides: useful precursors for N-modifications in common organic solvents *Carbohydr. Res.* 343 (2008) 2576-2582.
- [4] P. Sahariah, B.S. Snorraddottir, M.A. Hjalmarsdottir, O.E. Sigurjonsson, M. Masson, Experimental design for determining quantitative structure activity relationship for antibacterial chitosan derivatives, *J. Mater. Chem. B* 4(27) (2016) 4762-4770.

Design of genus-specific probes and primers for detection and identification of viral DNA in environmental samples using next-generation sequencing method

Andrey Ayginin^{1,2,*}, Anna Speranskaya^{1,3}, Alina Matsvai^{1,2}, Vladimir Dedkov¹, Marina Safonova¹, Ekaterina Pimkina¹, Dmitry Shagin¹, Mikhail Markelov¹, German Shipulin¹, Kamil Khafizov^{1,2}

¹ Central Research Institute of Epidemiology, Moscow, Russia

² Moscow Institute of Physics and Technology, Dolgoprudny, Russia

³ Lomonosov Moscow State University, Moscow, Russia

*aiginin@phystech.edu

Abstract. To date, the role of migratory birds in the circulation of a wide range of viral pathogens is well studied. Analysis of bird's seasonal migration routes and their viromes allows to predict the future directions and timing of pathogens transport, and may help to prevent infections spreading. Metagenomics techniques are widely used for similar studies, such as gut microbiota studies. However, it is impossible to create universal oligonucleotides to target all known viruses due to the viral genome diversity and their variability, whereas whole-genome sequencing is still expensive and relatively low-sensitive for such goals. In this work, we presented a genus-specific oligonucleotide panel for target enrichment of viral nucleic acids in different samples and possibility of its application for virus detection in samples collected from migratory birds. Our panel has been tested using a number of collected samples and has demonstrated superior efficiency in pathogen detection and identification.

Keywords: viromes; next generation sequencing; bioinformatics; primer design; metagenomics.

1 Introduction

Viruses, whose spreading is associated with migratory birds, have always attracted a close attention of specialists with respect to the possibility of their transport to considerable distances. Such migrations are tightly related to the emergence of epidemic and epizootic outbreaks as well as formation and activation of natural foci of viral infections. A significant number of human- and domestic animal-pathogenic viruses, such as severe acute respiratory syndrome virus (SARS), Influenza virus, West Nile virus, birds coronaviruses etc., are known to be transferred by migratory birds. The

asymptomatic carriage of viruses, which is referred to some features of bird metabolism and the adaptive capabilities of their immune system, produces the conditions for coevolution and the emergence of reassortants and recombinant strains of viruses. Due to this fact, the majority of widely-used methods of molecular diagnostics, e.g. those that employ polymerase chain reaction (PCR), are not quite suitable for identification of this type of viruses, because they are usually designed for detection of highly-conserved regions in the genomes. Moreover, a preliminary hypothesis of the presence of the certain viruses is required for performing the PCR test. If not, the process of the pathogen identification may take a significant amount of time, which can be crucial for prevention and control measures of the infection.

DNA barcoding is a method that uses a short part of the organism's genome (barcode) to identify whether it belongs to a particular specie by using massive parallel sequencing technologies. This method has been developed for studying bacterial communities, but today it is widely used for different tasks, including detection of food adulteration [1], marine communities' diets studying [2], biofuel analysis[3], etc. Unlike other taxa, viruses lack a shared universal phylogenetic marker (e.g., 16S for bacteria, Cytochrome C oxidase for birds and mammals, Internal transcribed spacer for fungi or plants), which makes it impossible to design the universal primer pair to amplify diverse viral sequences and to differentiate them. Furthermore, the taxonomy of viruses more clearly represents the signs of diseases they cause rather than a genetic similarity. This fact complicates the barcode selection even for one genus (for example, mammarenavirus that may be spitted into three or more subgenera, based on the genetic similarity and the area of distribution), let alone higher taxa. However, metagenomics still allows to detect different viral pathogens using shotgun sequencing. Though, this technique is quite expensive and is not suitable for screening of a big number of samples. Recently, Briese et al. [4] have developed a virome capture sequencing platform for vertebrate viruses (VirCapSeq-VERT), which consist of ~2 million biotinylated probes for target enrichment of viral nucleic acids for increasing the sensitivity of sequence-based virus detection and their characterization. The described method allows the detection of a large number of viruses, including the novel ones, however the overall cost of sequencing per sample still remains rather high.

Other method of enrichment is based on the target amplification of DNA region using specific primer pair in PCR. To date, a significant number of such primers have been described [5-7]. However, most of them have been designed for detection of certain virus species. Moreover, it is impossible to use them in the multiplex reaction due to different annealing temperatures and nonspecific amplification. Due to this fact, a number of PCR experiments that equals to the number of primer pairs is necessary to be carried out.

In this work, we present a method for designing oligonucleotide panels for targeted viral nucleic acids enrichment. The main goal of this research was the development of the algorithm that allows designing of the minimal number of oligonucleotides to cover the maximal number of diverse viral taxa. This method has been applied to design genus-specific biotinylated probes and primer pairs for target enrichment of viral nucleic acids. With both methods, the superior increase in viral genome coverage has been shown.

2 Materials and methods

The algorithm for the design of both genus-specific probes and primers has been developed and implemented to create the universal enrichment panels. Some restriction parameters were applied to the positions and the structures of oligonucleotides to permit the carrying out of enrichment reaction in one tube.

2.1 Database

The validated reference viral nucleic acids sequences are required for using the algorithm. One of the sources of such sequences is an open-source database “The Virus Pathogen Database and Analysis Resource (ViPR)” [8] which was developed in 2011. The most valuable advantage of this database is a high authenticity of the nucleic acids sequences, due to the data curation and management by the experts in virology and bioinformatics. This source was used to retrieve the sequences of the polymerase genes of the target viral genera (in some cases the different genes were used). The sequences were combined to the corresponding fasta files. The obtained data was filtered by record length (≥ 400 b.p), quality and intra-genus similarity (totally dissimilar to average genus sequence records were removed).

2.2 Common part of the algorithm

Genus-specific fasta files were aligned using ClustalW [9]. If the consensus subsequence with a specified length (20 for primer design; 40 for probe design) with 4 or less ambiguous position with minor nucleotide frequency $\geq 10\%$ was not identified, the initial fasta file was iteratively clustered with cd-hit [10] with the decreasing threshold tr . The clusters obtained during one step were aligned independently with ClustalW in order to identify consensus subsequence (-es) with described earlier properties for all of the subsets, which must represent $\geq 90\%$ different species inside the genus if combined together. Finally, one or more aligned fasta files were obtained for every genus.

Sliding window with the specified range of lengths was used for common subsequence extraction in multiple sequence alignment. The window “slides” from 5' to 3'-end of the alignment with a step equal to 1 nucleotide and identifies all subsequences, which meet the following criteria:

proportion of ambiguous positions (P_{amb}) $\leq 20\%$;

proportion of unique species, which share the subsequence and do not contain gaps (P_{sh}) $\geq 50\%$;

GC content of the consensus sequence (Pgc) is from 35% to 65%;
 absence of self-complementary regions;
 absence of homodimers formation;
 absence of formation of heterodimers with the previously selected oligonucleotides.

2.3 Probe design

The specificity of the extracted subsequences was evaluated using nucleotide blast [11] against the NCBI non-redundant (nr) nucleotide database and the reference viral database. The following parameters were calculated:

- Target_species – proportion of unique virus species related to target genus with alignment score ≥ 30 ;
- Non_target_species – number of virus species related to untargeted genus with alignment score ≥ 30 ;
- Median_score – median alignment score for target genus;
- Names_list – list of unique virus species related to target genus;
- Comp_to_earlier – maximal melting temperature for the heterodimer formation with earlier identified probes;
- Non_viral_species – number of non-viral species with alignment score ≥ 30 .

The remaining probe candidates were sorted with the following priority: descending by Target_species, then descending by Median_score, then ascending by Non_viral_species, and ascending by Non_target_species. The best sequence was selected as a genus probe. If it covered more than 90% of species inside the target genus or cluster, the algorithm was stopped. Else, the covered species were removed from the database and probe selection algorithm was reused with the restricted data. This step was repeated until total coverage of cluster or genus reaches 90% or more.

The following genera of viruses were chosen for the probes panel design (Table 1).

Genus
Alphacoronavirus
Betacoronavirus
Alphavirus
Lyssavirus
Flavivirus
Hantavirus

Phlebovirus
Mammarenavirus

Table 1. List of reference genera for the probes design.

2.4 Primer design

Some restrictions were applied to the primer design algorithm. First of all, total amplicon length between two subsequences was fixed between 200 and 400 b.p. to obtain compatibility of the panel with the most popular sequencing platforms, such as, for instance, Illumina MiSeq or Ion S5. Two subsequences were considered as a pair if both sequences cover more than 90% of species related to target genus or cluster and shared more than 90% of common species. The pairs were then filtered with respect to their annealing temperature ($50C \leq Ta \leq 55C$). Then blast against the nr database and the reference virus sequences database was used to check the specificity of the selected pairs. Nonspecific candidates were removed. The possibility of heterodimers formation between the sequences inside the pair and between the sequences and the previously selected primers was calculated using Primer3 software [12]. Then the same parameters as those described in the probes design section were calculated and the pairs were sorted accordingly. The best primer pair was selected as a “genus pair”.

The primer pairs for the reference genera of viruses designed using the developed method are shown in Table 2.

Genus	Primer pairs	Control sample species	Abbreviation
Alphavirus	Alpha-F/R	-	-
Ebolavirus	Ebola(2)-F/R	Ebola virus	EBO
Coltivirus	Colti-F/R	-	-
Cuevavirus	Cueva-F/R	-	-
Ledantevirus	Ledante-F/R	-	-
Lyssavirus	Lyssa(1)-F/R Lyssa(2)-F/R	Rabies lyssa-virus	RLV
Mammarenavirus	Mammarena-F/R	Unidentified lassavirus	LASS
Marburgvirus	Marburg(1)-F/R Marburg(2)-F/R	Marburg virus	MAR
Orbivirus	Orbi-F/R	-	-
Orthobunyavirus	Orthobunya-F/R	Unidentified orthobunyavirus	ORTHB

Betacoronavirus	Corona(b1)-F/R Corona(b2)-F/R	Unidentified coronavirus	COR
Thogotovirus	Thogoto-F/R	-	-
Gammacoronavirus	Corona(g1)-F/R	Unidentified coronavirus	COR
Alphacoronavirus	Corona(a1)-F/R	Unidentified coronavirus	COR
Hantavirus	Hanta-F/R	Dobrava-Belgrade orthohantavirus	DBO
Phlebovirus	Phlebo-F/R	-	-
Cardiovirus	Cardio-F/R	-	-
Orthoreovirus	Orthoreo-F/R	-	-
Parechovirus	Parecho-F/R	-	-
Seadornavirus	Seadorna-F/R	-	-
Henipavirus	Henipa-F/R	-	-
Nairovirus	Nairo-F/R	-	-
Flavivirus	Flavi-F/R	Yellow fever virus	YFV
		Japanese encephalitis virus	JFV

Table 2. List of reference genera the for primers design.

2.5 Testing

The enrichment efficiency of the designed probes were tested using the reference samples which contained the RNA of the virus species: sample “A” contained Tick-borne Encephalitis virus (TBEV, flavivirus), sample “B” – Yellow fever virus (YFV, flavivirus) and sample “C” – Rabies lyssavirus (RLV, lyssavirus). The “REVERTA-L” kit (AmpliSens) was used for reverse transcription reaction to obtain the cDNA.

For testing the probes pool, the library preparation and target enrichment by the hybridization with the probe panel was carried out using the in-house protocols. The NGS libraries after hybridization as well as the same libraries without target enrichment were sequenced on the Illumina MiSeq. The obtained reads were filtered and aligned to the reference viral genomes in order to calculate the coverage of the target loci.

The primer panel was tested with the reference set of reference samples (RNA feces), which contained the RNA of the virus species previously characterized by Illumina sequencing (see Table 2).

The primer panel was tested by carrying out simplex and multiplex PCR with the samples. Firstly, the primer pairs were tested independently in simplex PCR. For the panel validation, the primers were mixed and tested together in multiplex PCR with the same samples. The re-amplification reaction with the PCR products was carried out with the individual primer pairs to control the amplicon length. In all cases the amplicons were detected using electrophoresis.

3 Results

The results of the experiments with control reference samples and set of probes showed an appropriate enrichment efficiency (Fig. 1). The coverage of the YFV genome has increased from approximately 1-2 reads for all regions of genome for the pure sample to ~ 60 for target region of the enriched sample. For the TBE, it has increased from ~ 1-2 reads for all regions to ~ 15 for target region; for RLV – from 2 reads to ~ 80 reads.

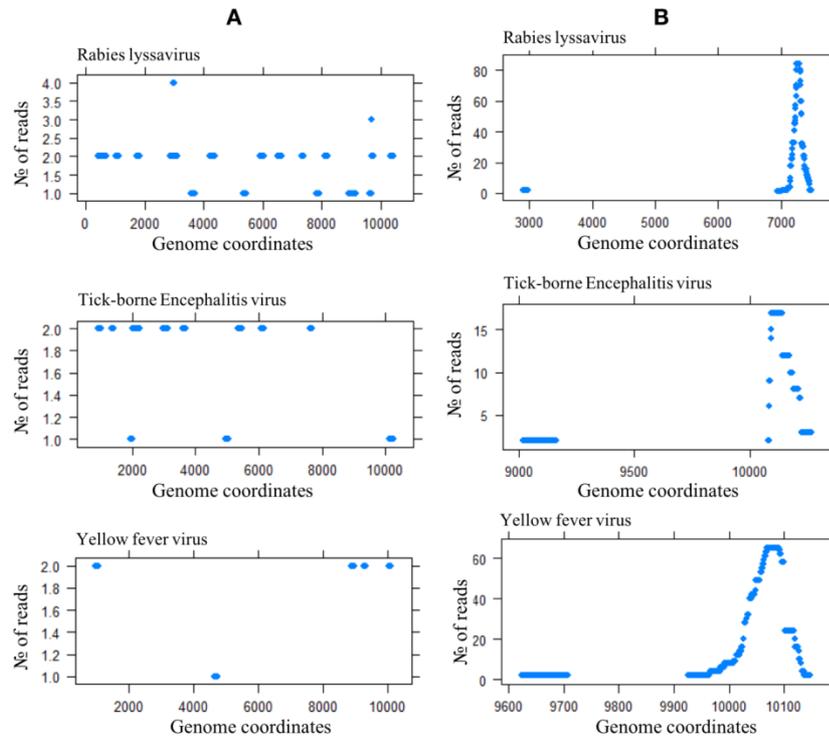


Fig. 1. The coverage of the reference genomes. **A** – without enrichment **B** – after enrichment

The primer panel was tested with a reference set of samples: The results of the individual testing are shown in Figs. 2 and 3. In all cases the products of the specified range of lengths were obtained.

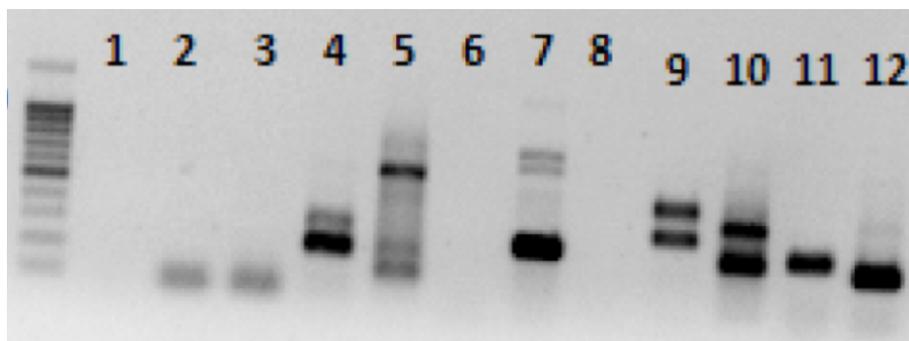


Fig. 2. Individual primer pairs testings, part 1. 1 – Negative control, 2 – MAR, primers Marburg(1)-F/R, 3 – MAR, primers Marburg(2)-F/R, 4 – YFV, 5 – JEV, 6-9 – COR (primers Corona-a1, Corona-b1, Corona-b2, Corona-g1 respectively), 10 – DBO, 11 – LASS, 12 – ORTHB.

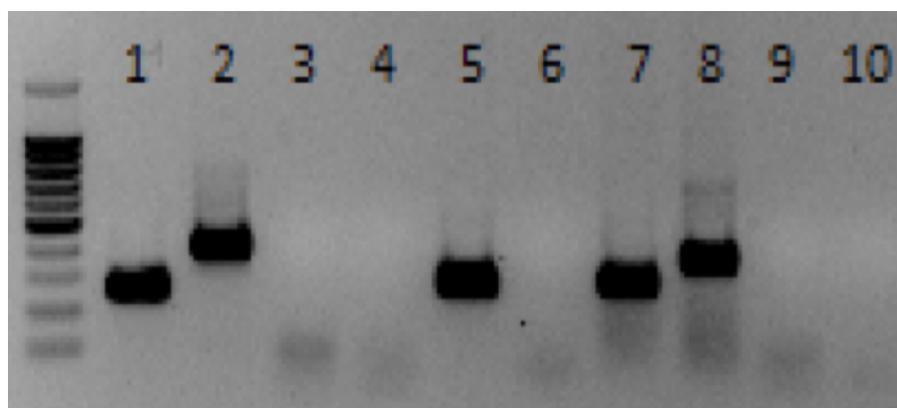


Fig. 3. Individual primer pairs testings, part 2. 1 – MAR, primers Marburg(1)-F/R, 2 – MAR, primers Marburg(2)-F/R, 3 – negative control for primers Marburg(1)-F/R, 4 – negative control for primers Marburg(2)-F/R, 5 – EBO, 6 – negative control for primers EBO, 7 – RLV, primers Lyssa(1)-F/R, 8 – RLV, primers Lyssa(1)-F/R, 9 – negative control for primers RLV, primers Lyssa(1)-F/R, 10 – negative control for primers RLV, primers Lyssa(1)-F/R.

The multiplex system was tested with the same set of viruses and the PCR products were re-amplified using the genus-specific primer pairs. In all cases, unspecific amplification was observed. However, re-amplification reactions with the genus-specific primers have shown the target products presence. The multiplex system was tested with the three bats samples infected by several known viruses (sample 1 – betacoronavirus, sample 2 – betacoronavirus, sample 3 – orthoreovirus). Reamplification of the PCR products was carried out with the genus-specific primers (Fig. 4). The obtained results clearly demonstrate the presence of the target products.

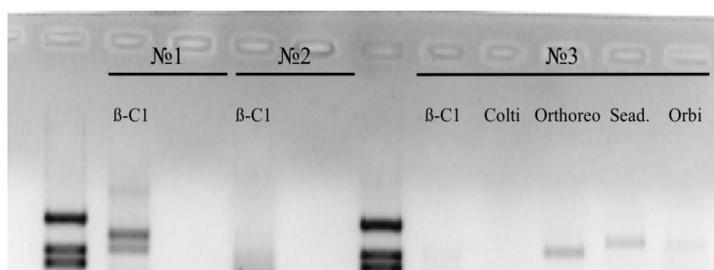


Fig. 4. Reamplification of the multiplex PCR product; β -C1 - betacoronavirus-specific primers; Colti - coltivirus-specific primers; Orthoreo - orthoreovirus-specific primers; Sead. - Seadornavirus-specific primers; Orbi - orbivirus-specific primers

4 Discussion

Virome analysis of migratory birds attracted a close attention of the specialists due to the possibility of an accurate prediction of regions that might be affected by infections. It is especially important for viruses, which are known to persist in birds and share high genetic diversity, and which can be potentially dangerous for humans and animals. Due to these facts, the development of rapid test-systems is necessary for effective regulation of epidemic situation. In this work, we have presented a method of oligonucleotide design for viral nucleic acids enrichment. We have used our method to design the DNA-probes and primer panels, which have shown high efficiency in detection and identification of multiple reference viruses using NGS methods. The described method may be recommended to researchers, who need to investigate diverse viromes in the various sample types. In the future, we are planning to expand the panel to target more diverse viruses and test it with the nature bird samples from diverse regions.

This work was supported by RSF grant N. 17-74-20096.

References

1. Parvathy, V.A., et al., Detection of plant-based adulterants in turmeric powder using DNA barcoding. *Pharm Biol*, 2015. 53(12): p. 1774-9.
2. Berry, T.E., et al., DNA metabarcoding for diet analysis and biodiversity: A case study using the endangered Australian sea lion (*Neophoca cinerea*). *Ecol Evol*, 2017. 7(14): p. 5435-5453.
3. Jaenicke, S., et al., Comparative and joint analysis of two metagenomic datasets from a biogas fermenter obtained by 454-pyrosequencing. *PLoS One*, 2011. 6(1): p. e14519.
4. Briese, T., et al., Virome Capture Sequencing Enables Sensitive Viral Diagnosis and Comprehensive Virome Analysis. *MBio*, 2015. 6(5): p. e01491-15.
5. Zheng, L., et al., Novel genus-specific broad range primers for the detection of furoviruses, hordeiviruses and rymoviruses and their application in field surveys in South-East Australia. *J Virol Methods*, 2015. 214: p. 1-9.
6. Choi, S.K., et al., RT-PCR detection and identification of three species of cucumoviruses with a genus-specific single pair of primers. *J Virol Methods*, 1999. 83(1-2): p. 67-73.
7. Pfeffer, M., et al., Genus-specific detection of alphaviruses by a semi-nested reverse transcription-polymerase chain reaction. *Am J Trop Med Hyg*, 1997. 57(6): p. 709-18.

8. Pickett, B.E., et al., ViPR: an open bioinformatics database and analysis resource for virology research. *Nucleic Acids Res*, 2012. 40(Database issue): p. D593-8.
9. Larkin, M.A., et al., Clustal W and Clustal X version 2.0. *Bioinformatics*, 2007. 23(21): p. 2947-8.
10. Li, W. and A. Godzik, Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, 2006. 22(13): p. 1658-9.
11. Altschul, S.F., et al., Basic local alignment search tool. *J Mol Biol*, 1990. 215(3): p. 403-10.
12. Untergasser, A., et al., Primer3--new capabilities and interfaces. *Nucleic Acids Res*, 2012. 40(15): p. e115.

PIKA VIRUS: A METAGENOMICS TOOL FOR VIRAL COMMUNITY ANALYSIS

Andrea Rubio¹, Sara Monzon¹, Pedro J. Sola-Campoy^{1,2}, Isabel Cuesta¹

¹ Bioinformatics Unit, BU-ISCIH, Institute of Health Carlos III, Majadahonda, Spain
² Reference and Research Laboratory for Antibiotic Resistance, CNM, ISCIH, Majadahonda, Madrid.

smonzon@sciii.es

Abstract.

In recent years, metagenomics has emerged as a powerful means for pathogen identification and discovery. Although many metagenomics pipelines have been developed to date, they do not currently present enough data for reliable diagnosis, as vital pieces of information are not provided within their results report. On the other hand, the report is usually cryptic and difficult to follow and understand. We have developed an open-source new bioinformatics workflow for metagenomics analysis which provides extensive and rigorous information presented in a user-friendly web format. PikaVirus implements a new method combining assembly and mapping information for more accurate virus identification from shotgun metagenomics data. All the project code can be downloaded from <https://github.com/BU-ISCIH/PikaVirus>. There is also a wiki page where all the information needed for running the pipeline can be found at <https://github.com/BU-ISCIH/PikaVirus/wiki>.

Keywords: metagenomics, virus, mapping, assembly, report

1 Introduction

Metagenomics, the study of the genetic material of a given microbial community, existed before the Next generation Sequencing (NGS) breakthrough, although its main goal was restricted to establishing the composition of the microbial communities by

sequencing rRNA [1], it wasn't until the NGS techniques took over that metagenomics started having an important role in pathogen detection.

The qualitative composition of the sample can be established today by two different approaches: either by directly comparing the reads with a database (VirusTAP[2]), or by the assembly of the reads into longer sequences, commonly known as contigs, before comparing them with the reference (Pathosphere [3]). The estimation of each organism abundance is also possible, but it is highly dependent on an unbiased sample preparation in order to be representative of the original sample composition [4].

Our aim was the development of a new metagenomics workflow based on both mapping against a reference genome database and assembly approaches and combining blast information with coverage genome data for the detection and determination of diverse pathogens, especially virus, in shotgun NGS data.

2 Materials and methods

2.1 Software development

PikaVirus has been primarily implemented in bash, R v3.2 and python 2.7, using JavaScript, CSS3 and HTML5 for the web-based report. The project code is open source and it is available at <https://github.com/BU-ISCI/III/PikaVirus>.

2.2 Respiratory samples

Four nasopharyngeal aspirates sequenced in Spanish National Center for Microbiology where used for PikaVirus testing. Reads where obtained from Illumina Nextseq machine. Samples were also analyzed in Pathosphere for results comparison.

3 Results

3.1 PikaVirus workflow

The general workflow of PikaVirus starts with the raw pair-end (PE) reads, which are pre-processed to ensure minimum quality requirements. After that, human reads are filtered out by mapping against the human. Once only non-host reads remain, they are all mapped against Virus, Bacteria, Fungi, Protozoa and Invertebrate reference genome databases. Once the reads are binned, they are assembled into contigs which are then blasted with their respective reference database to find their best match. In parallel, the genome coverage of each represented genome is calculated. Reads that fail to be binned into one of the groups are independently assembled and blasted against the virus reference genome database. Finally, all the results are merged and presented in web format.

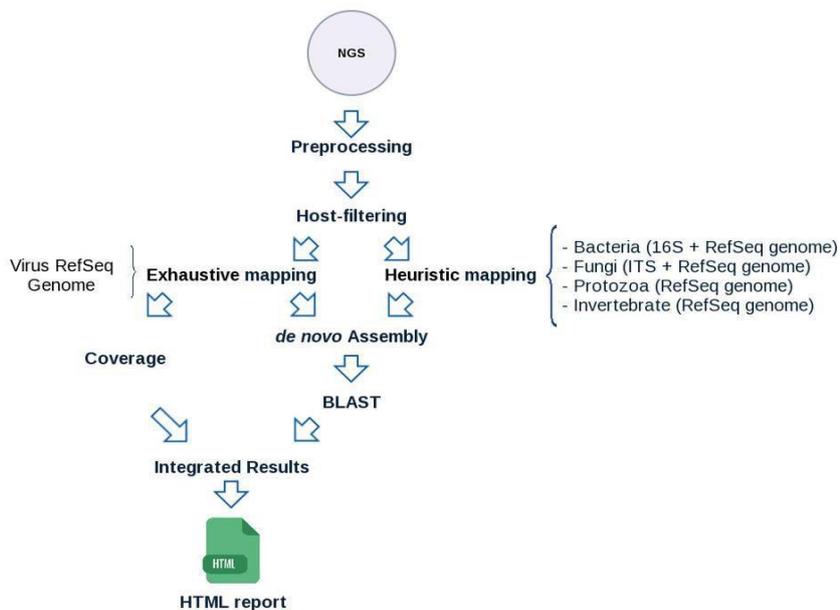


Fig. 1. PikaVirus workflow scheme.

3.2 Friendly report web-based for helping researcher interpretation

One of the goals of PikaVirus was to provide the researcher with navigation-friendly and easy to understand output analysis. To that account, the results of the analysis are presented in a web page with the following layout:

- **Summary.** The summary section contains an overview of the sample's results so that the most represented organisms in each sample can be viewed at a glance.

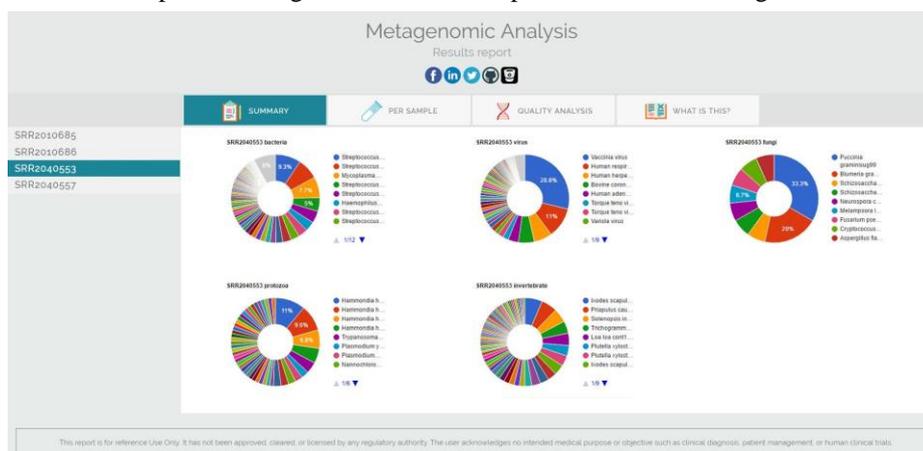


Fig. 2. HTML report summary section. A pie chart is presented for each organism group and for each sample containing the best blast hit of each contig in the sample.

- **Per sample.** This section contains in depth information about the samples. The blast and coverage analysis are merged for each contig, organized by sample and organism; and shown as a single line of a table.
- **Read quality analysis.** This section organizes trimming and read quality control data in a practical way, so they can be easily consulted.
- **What Is This?** One of the main goals of generating the analysis report in web format is to provide an intuitive, user-friendly report for the wet-lab researcher, not used to dealing with plain text or log files. This section explains how to read each section of the results page and how the data contained in them is calculated.

Metagenomic Analysis
Results report

f in t y

SUMMARY PER SAMPLE QUALITY ANALYSIS WHAT IS THIS?

Reference Id	Reference name	Contig Id	% of identical matches	Alignment length	Number of mismatches	Number of gap openings	Start of alignment in query	End of alignment in query
AC_0000071	Human adenovirus 2, complete genome	NODE_206_length_337_cov_8.08779	99.12	227	1	1	66	291
AC_0000081	Human adenovirus 5, complete genome	NODE_206_length_337_cov_8.08779	99.12	226	0	1	66	291
AC_0000101	Simian adenovirus 21, complete genome	NODE_345_length_289_cov_3.17949	91.02	258	23	0	1	258
AC_0000101	Simian adenovirus 21, complete genome	NODE_345_length_215_cov_2.625	93.85	179	7	2	40	214
AC_0000171	Human adenovirus type 1, complete genome	NODE_206_length_337_cov_8.08779	99.12	227	1	1	66	291
AC_0000181	Human adenovirus type 7, complete genome	NODE_228_length_302_cov_2.2996	100	302	0	0	1	302
AC_0000181	Human adenovirus type 7, complete genome	NODE_345_length_289_cov_3.17949	99.95	289	1	0	1	289
AC_0000181	Human adenovirus type 7, complete genome	NODE_250_length_385_cov_1.82869	96.88	241	8	0	45	285
AC_0000181	Human adenovirus type 7, complete genome	NODE_130_length_337_cov_3.7473	98.42	317	5	0	1	317
AC_0000181	Human adenovirus type 7, complete genome	NODE_398_length_242_cov_12.1237	98.46	130	2	0	112	241
AC_0000181	Human adenovirus type 7, complete genome	NODE_345_length_215_cov_2.625	92.61	230	2	2	1	215
AC_0000181	Human adenovirus type 7, complete genome	NODE_346_length_215_cov_2.1875	99.07	215	2	0	1	215

This report is for reference Use Only. It has not been approved, cleared, or licensed by any regulatory authority. The user acknowledges no intended medical purpose or objective such as clinical diagnosis, patient management, or human clinical trials.

Fig. 3. HTML report per sample section. A table with blast and coverage information per sample and per organism is reported.

3.3 Mapping and coverage for a more accurate virus identification

The comparison with Pathosphere provided the expected insight about the genome coverage as metric for reliable clinical diagnostic as shown in Table 1. Both PikaVirus and Pathosphere found several Human adenovirus (HAdV) in a nasopharyngeal sample analyzed, for example. In both cases the identity percentage of all the H. adenovirus types ranged between 70% and 100% approx. With this data alone, it is not possible to determine which of the HAdV types is present in the sample. However, adding the genome coverage information obtained from mapping, we are able to determine that H. adenovirus 2 genome is covered in almost its totality (99.7%), while the others are only 86%-88% covered. This data allows us to deduce the identity reported by blast is caused by the high homology between all the HAdV

Accession number*	Genome	PikaVirus			Pathosphere	
		Min identity (%)	Max identity (%)	Co-verage (%)	Min identity (%)	Max identity (%)
AC_000007.1	<i>Human adenovirus 2</i>	99.3	100	97.5	77.3	100
AC_000008.1	<i>Human adenovirus 5</i>	94.6	100	86.5	-	-
AC_000017.1	<i>Human adenovirus 1</i>	92.1	100	87.7	97.87	100
NC_003216.1	<i>Human adenovirus 6</i>	-	-	-	81.03	99.37

Table 1. Comparison between PikaVirus and Pathosphere results. Green shows coverage between 90-100%, yellow between 50-89% and red between 0-49%.

4 Discussion

The number of metagenomics projects has greatly increased the last decade, mainly due to the overwhelming improvement metagenomics approaches present over classical molecular ones for pathogen identification and study[5].

Of the various bioinformatics tools available to date, we found none handle and present data in a way that we find satisfying, either because they perform black-box analysis of the data or because they present cryptic and incomplete results.

For that reason, we developed PikaVirus, a metagenomics tool for viral community analysis. PikaVirus provides an in depth and user-friendly report and is completely open source.

Most metagenomics software use assembly as the main approach in their analysis. Preliminary results of our approximation combining mapping with assembly and adding coverage genome approach show an increased specificity. However, a comparison using better datasets is needed to establish the exact specificity.

References

1. Olsen, G. J., Lane, D. J., Giovannoni, S. J., Pace, N. R. & Stahl, D.A.: Microbial ecology and evolution: a ribosomal RNA approach. *Annu. Rev. Microbiol.* 40, 337–365 (1986)
2. Yamashita, A., Sekizuka, T., Kuroda, M.: VirusTAP: Viral Genome-Targeted Assembly Pipeline. *Front. Microbiol.* 7, 32 (2016). doi:10.3389/fmicb.2016.00032
3. Kilianski, A., Carcel, P., Yao, S., Roth, P., Schulte, J., Donarum, G.B., Fochler, E.T., Hill, J.M., Liem, A.T., Wiley, M.R., Ladner, J.T., Pfeffer, B.P., Elliot, O., Petrosov, A., Jima, D.D., Vallard, T.G., Melendrez, M.C., Skowronski, E., Quan, P.-L., Lipkin, W.I., Gibbons, H.S., Hirschberg, D.L., Palacios, G.F.,

- Rosenzweig, C.N.: Pathosphere.org: pathogen detection and characterization through a web-based, open source informatics platform. *BMC Bioinformatics*. 16, 416 (2015). doi:10.1186/s12859-015-0840-5
4. Head, S.R., Kiyomi Komori, H., LaMere, S.A., Whisenant, T., Van Nieuwerburgh, F., Salomon, D.R., Ordoukhanian, P.: Library construction for next-generation sequencing: Overviews and challenges. *Biotechniques*. 56, 61–77 (2014). doi:10.2144/000114133
 5. Reddy, T.B.K., Thomas, A.D., Stamatis, D., Bertsch, J., Isbandi, M., Jansson, J., Mallajosyula, J., Pagani, I., Lobos, E.A., Kyripides, N.C.: The Genomes OnLine Database (GOLD) v.5: a metadata management system based on a four level (meta)genome project classification. *Nucleic Acids Res.* 43, D1099–D1106 (2015). doi:10.1093/nar/gku950

Gene Expression-based Cancer Classification with Handling the Class Imbalance Problem

Sadam Al-Azani and Emad Ramadan

Department of Information and Computer Science,
King Fahd University of Petroleum and Minerals,
Dhahran, Kingdom of Saudi Arabia
alazani, eramadan@kfupm.edu.sa

Abstract. Cancer is a leading cause of death and majority of cancer cases are diagnosed in the late stages of cancer by using conventional methods. The gene expression microarray technology is applied to detect and diagnose most types of cancers in their early stages. In this paper, we address the problem of classifying cancer based on gene expression with handling the class imbalance problem. The synthetic minority over-sampling technique (SMOTE) is utilized to overcome this problem by adding synthetic samples. Another common issue related to the gene expression dataset that is addressed in this paper is the curse of dimensionality. This problem is addressed through applying Chi-Square and Information Gain feature selection techniques. After applying these techniques individually, we proposed a method to select the most significant genes by combining those two techniques (CHIS& IG) and investigated the effect of these techniques individually and in combination. Three benchmarking bio-medical datasets(Leukemia-subtypes, Leukemia-ALLAML and Colon) are used. As classification methods, individual and ensemble learning methods are employed and compared. As individual classifiers MLP and SMO-SVM are used. As ensemble classifiers, we used Random forests classifier and proposed a majority voting based ensemble of k-NN classifiers. The experimental results reveal that the SMOTE technique results in improving the results in the most cases. Additionally, the performance of the combined feature selection techniques outperforms individual techniques in many cases. The obtained results compare favourably with the literature.

Keywords: Cancer detection and diagnosis, Gene expression, Feature selection, Class imbalance, ensemble classifiers

1 Introduction

A gene is a small piece of genetic material written in a code called DNA. Each gene has within it a set of instructions for making molecules that organisms need to survive. Genes themselves cannot be used by an organism. Instead, they must be turned into a gene product. Gene expression is the process by which the information contained within a gene becomes a useful product.

Clinical and pathological information might be incomplete or misleading for cancer detection and diagnosis [1]. Cancer detection and diagnosis can be objective and highly accurate using microarrays, which have the ability to support clinicians with the information to choose the most appropriate forms of treatment [2]. Considering detection and diagnosis cancer using gene expressions as classification problem is traced back to the work of [3] which classify acute leukemias of human as a test case.

The major challenge in microarray data is the curse of dimensionality such that they include large number of genes with small number of samples. To alleviate this limitation, researchers proposed several feature selection techniques that could select the most significant genes and investigated several statistical and machine learning classification methods.

Another critical issue regarding microarray datasets is the class imbalance problem in which there is significant variation in the samples per a class. This issue, to our knowledge, has not addressed in the case of classifying cancers using gene expression data. Therefore, one of the purpose of this research is to address this issue using Synthetic minority over-sampling technique (SMOTE) [4]. SMOTE is an over-sampling algorithm aims at overcoming the problem of imbalanced dataset by over-sampling the minority class [4]. This is conducted by generating artificial samples instead of over-sampling with replacement. For other biomedical applications, some important techniques for class prediction of high-dimensional imbalanced data was reviewed and classified by [5].

In this research: an approach to classify the cancers based on gene expression datasets with handling the class imbalance problem is proposed. Three publicly available gene expression microarray datasets which suffer the class imbalance problem and the curse of dimensionality are used. They are binary and multiple classes. The effectiveness of the combining feature selection techniques was investigated. Furthermore, multilayer perceptron (MLP), sequential minimum optimization based support vector machine (SMO-SVM) and Random forests classifiers are applied and compared. In addition, a homogeneous ensemble of k-NN classifiers based on the majority voting method is proposed in which the base k-NN models are built through the learning of different values of k. The rest of this paper is organized as follows. In Section 2 we reviewed the most related works while in Section 3 we present the experiments and results. Section 4 concludes this research.

2 Related Works

Lee et al. [6] used neural network-based finite impulse response extreme learning machine (FIR-ELM) to classify leukemia and colon tumor. The proposed method (FIR-ELM) performs better than other classifiers for leukemia dataset. For colon dataset, SVM performs better than FIR-ELM. Lotfi and Keshavarz [7] presented an approach called PCA-BEL based on principle component analysis and brain emotional learning network for classification gene-expression microarray data. The pros of using BEL is that its low computational complexity.

Rathore, Iftikhar, and Hussain [8] used two feature selection techniques in sequence. The first feature selection technique, Chi-square takes the whole dataset and selects the discriminative gene subset which then used as an input to the second feature selection technique, mRMR, which selects the most discriminative gene subset among them. They reported that the proposed technique achieve classification rates and perform better than the individual techniques.

Another approach that is utilized the nature of variations in gene expressions to classify colon gene samples into normal and malignant classes is conducted by Rathore, Hussain, and Khan [9]. They presented a majority voting based ensemble of SVM. Experimental results reveal that the ensemble classifier improve the results comparing with the single classification techniques.

Bouazza et al. [10] reported that SNR (Signal-to-Noise Ratio) feature selection technique is the most trusted technique selecting the genes from three different datasets when compared with Fisher, T-Statistics and ReliefF. Banka and Dara [11] proposed a Hamming distance based binary Particle Swarm Optimization (HDBPSO) algorithm to select the most significant gene subsets.

Simjanoska, Bogdanova, and Popeska [12] analyzed the gene expression to classify the colon carcinogenic tissue. The Illumina HumanRef-8 v3.0 Expression BeadChip microarray technology was utilized to do the gene expression profiling which contains 26 colorectal tumors and 26 colorectal healthy tissues. An original methodology contains several steps was developed for biomarkers detection which include data preprocessing, statistical analysis, modeling the a priori probability for all significant genes. It was reported that Bayes theorem performs better than SVM and BDT. These findings were justified due to the realistic modeling of the priori probability of Bayes' theorem. However, such method is somewhat complicated. The priori probability model generated in [12] is then employed in Bogdanova, Simjanoska, and Popeska [13] by using gene expression with the Affymetrix Human Genome U133 Plus 2.0 Array, the results revealed poor distinctive capability of the biomarker genes. That means the priori probability model is platform dependent. This finding confirms what was concluded in Wong, Loh, and Eisenhaber [14] where they stated each platform requires different statistical treatment [14]. Simjanoska, Madevska Bogdanova, and Popeska [15] generated a statistical approach of gene expression values obtained from Affymetrix using the similar methodology of [12]. The findings revealed that an excellent results were achieved using Bayes' theorem when appropriate preprocessing methodology is followed. The results reported in [15] was then improved in Simjanoska and Bogdanova [16]. They proposed a filtering genes method to select the most essential biomarkers. It is called Leave-one-out method and is based on iterative Bayesian classification.

Tong et al. [17] introduced a genetic algorithm (GA) based ensemble SVM classifier constructed on gene pairs (GA-ESP). The base classification methods (SVMs) of the ensemble system are trained on various informative gene pairs. These gene pairs are selected by the top scoring pair (TSP) criterion. GA is then employed to select an optimized combination of base classifiers. The applicabil-

ity of the proposed approach was evaluated on several cancer datasets in both binary-class and multiple class datasets.

Cao et al. [18] presented a novel fast feature selection method based on multiple SVDD, called MSVDD-RF. Insignificant features are eliminated recursively. It is applied to multiple class microarray data to detect different cancers. Table 1 summarizes the reviewed studies in terms of the type of cancer, the used and proposed techniques, the features selection techniques and the used datasets.

In conclusion, none of the aforementioned studies has addressed the problem of class imbalance. Additionally, nearly all of those studies applied the feature selection techniques individually apart from that is the study of Rathore et al. [8]. The main difference is that they used two feature selection techniques in sequence (ChiS and mRMR) while we combined two techniques (ChiS and IG) in parallel. This is in addition to the difference in the methodologies. Furthermore, rare efforts have been conducted to employ ensemble learning methods.

Table 1. Related Works

Ref	Cancer Type	Technique	Feature Selection	Dataset
[6]	Lukemia, colon	FIR-ELM, MLP, ELM, SVM	FGFS	KentRidge
[7]	Lung, colon, breast	ANN: PCA-BEL	PCA	KentRidge
[8]	Colon	SVM	Chi-squar+mRMR	KentRidge, Notterman, E-GEOD-40966
[9]	Colon	SVM, ensemble of SVMs	chi-square, F-Score, PCA, mRMR	KentRidge, BioGPS, Notterman, E-GEOD-40966
[10]	Leukemia, prostate, colon	K-NN, SVM	Fisher, T-Statistics, SNR (Signal-to-Noise Ratio), ReliefF	KentRidge, GEO
[11]	Colon, lymphoma, leukemia	k-NN, BLR, Bayes Net, Neuroblastoma, SVM, MLP, J48, LMT, Random forests.	HDBPSO	
[12]	Colon	Bayes classifier, SVM, BDT		GEO
[13]	Colon	Bayes classifier		Affymetrix Human Genome
[16]	Colon	Bayes classifier	Low entropy filter, and Leave-one-out method	Affymetrix Human Genome
[17]	Breast, Colon, Leukemia, Prostate	GA, ensemble SVM	TCP	KentRidge
[18]	Colon, Lung, Leukemia	SVM, k-NN	MSVDD-RF	KentRidge, Leukemia3, Novartis

Table 2. Summary of the datasets

Dataset	Class	Gene	Training set	Test set
Leukemia-subtype	7 (BCR-ABL, E2A-PBX1, Hyperdiploid>50, MLL, T-ALL, TEL-AML1 and others)	12558	215 (9/ 18/ 42/ 14/ 28/ 52/ 52)	112 (6/ 9/ 22/ 6/ 15/ 27/ 27)
Leukemia-ALLAML	2 (ALL/AML)	7129	38 (27/11)	34 (20/14)
Colon	2 (Tumor/Normal)	2000	43 (28/15)	19 (12/7)

3 Experiments and Results

3.1 Gene Expression Microarray Datasets

Three publicly available gene expression microarray datasets, were used. All of these datasets suffer the problem of the class imbalance and the curse of dimensionality. The first one is Lymphoblastic Leukemia (Leukemia-subtype) dataset [19] and is a multiple class dataset. It is composed of seven subtypes (BCR-ABL, E2A-PBX1, Hyperdiploid>50, MLL, T-ALL, TEL-AML1 and others) and contains 12558 genes per sample. The second dataset is Leukemia-ALLAML [3] which is binary class (ALL and AML) and is composed of 7129 genes. The third dataset is colon tumor [20] which is a binary class (normal and malignant) and is composed of 2000 genes. Table 2 summarizes the datasets.

Each gene is normalized on the interval [0, 1].

3.2 SMOTE

For each dataset described in Table 2, another dataset is generated after applying SMOTE technique. SMOTE handle the problem of imbalanced dataset by over-sampling the minority classes [4]. Generating artificial examples by using SMOTE algorithm relies on the feature space similarities between existing minority examples. SMOTE considers the K-nearest neighbors of each example $x_i \in S_{min}$ such that S_{min} is the minority class and K is an integer number. It defines the K-nearest neighbors belongs to S_{min} with smallest Euclidean distance measure to x_i along the n-dimensions of feature space X . To generate synthetic samples, SMOTE computes the difference between the sample under consideration and its nearest neighbor and multiplies it by a random number between zero and one [21]. In the our experiments, we used k=5.

Applying SMOTE is performed for the training datasets because it is not reasonable to generate synthetic testing data and evaluating the models based on synthetic testing samples. After applying SMOTE the number training samples of each class in all created datasets become equal to the maximum number of training samples in the original ones. That means, the numbers of training samples per class in Leukemia-subtype, Leukemia-ALLAML and Colon become 52, 27 and 28, respectively. Table 2 also shows the majority and minority classes distributions the datasets.

3.3 Feature Selection

Two feature selection techniques namely, Chi-Square (ChiS) and Information Gain (IG) were investigated. Chi-S and IG rank the attributes based on their significance. Chi-Square evaluates attributes with respect to the class through computing the Chi-square statistic [22]. With IG feature selection technique, the worth of an attribute is evaluated through measuring the information gain with respect to the class. WEKA-3.6.13 [22] is used to implement ChiS and IG feature selection techniques. We adopted the cutoff threshold to zero because the default threshold did not reduce in the number of the genes, it just ranks them. We proposed a method to select the most significant genes by combining both ChiS and IG. We assumed that those genes that are selected by both techniques are more significant than those selected by the individual technique, so we selected the common genes selected by both techniques. Additionally, this generates less number of selected genes compared with the individual techniques. The number of full and selected genes of the datasets after applying feature selection techniques is shown in Table 3.

Table 3. Number of Genes in the Datasets before and after Applying Feature Selection Techniques

Technique	Leukemia-subtype	Leukemia-ALLAML	Colon
Full Genes	12558	7129	2000
ChiS	300	200	150
IG	300	200	150
ChiSIG	233	145	12
ChiSIG after applying SMOTE	255	163	111

3.4 Classification phase

MLP, SMO-SVM and Random forests classifiers are used and compared to classify the type of cancer. We investigated these classifiers with different parameters and structures implemented on WEKA-3.6.13 [22]. For MLP, different structures are investigated using different number of neurons in the hidden layers and learning rates. In the first structure, the number of hidden layer neuron is calculated by the taking the average of the number of input and the output, for example if the number of genes is 300 and the number of output (classes) is seven (number of classes) then the number of the neurons in the hidden layer is $(300 + 7)/2 = 154$. For learning rate we used the value of 0.3. Other structures of MLP were investigated with using values of 20, 50 and 80 as the number of neurons in the hidden layer with investigating different learning rates of 0.1, 0.3 and 0.5, with the possible combinations. All structures of MLP were investigated with Momentum of 0.2 with back propagation learning algorithm.

For SVM classifier, SMO-SVM algorithm was used with investigating two different kernel functions: PUK and Poly-kernel with the complexity parameter C of

one. Multi-class problem were solved using pairwise classification which is known as one-vs-one. SMO-SVM [23] is an algorithm used to speed up the training of SVM through breaking a very large quadratic programming (QP) optimization problem in SVM into a series of smallest possible QP problems. This in turn avoids using a time-consuming numerical QP optimization as an inner loop. Random forests combines decision tree predictors. Each tree relies on the values of a random vector which is sampled independently and all trees in the forest have the same distribution [24].

We proposed a homogeneous ensemble of k-NN classifiers to classify cancers by using the concept of majority voting of the predicted labels of the individual k-NN models. The majority voting algorithm combines the labels predicted by each of 1-NN, 3-NN and 5-NN models individually (1-NN means one nearest neighbour, 3-NN means three nearest neighbour, etc.). Then it assigns a label to a sample based on the majority votes. Fig. 1 illustrates the framework of the majority voting based ensemble of k-NN classifiers.

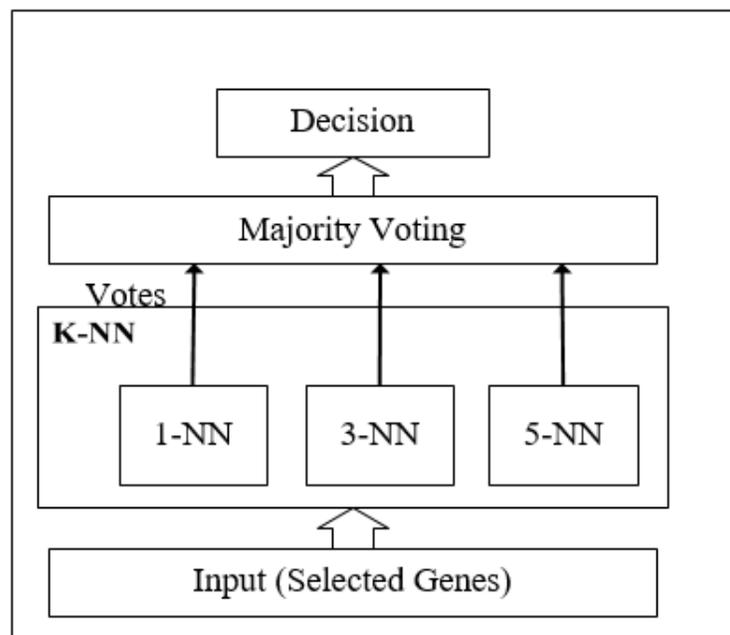


Fig. 1. The framework of the majority voting based ensemble of k-NN classifiers.

3.5 Results

Well-known performance metrics, namely precision, recall, F-measure and accuracy, are used in this work. These metrics require computing the number of positive samples classified correctly, called true positive (TP), the number of positive samples classified incorrectly, called false negative (FN), the number of negative samples classified correctly, called true negative (TN), and the number of negative samples classified incorrectly, called false positive (FP). The experimental results are summarized in Tables 4, 5 and 6 for MLP, SMO-SVM and Random forests, respectively. Each table show the results before and after applying SMOTE technique for each feature selection technique and is divided into three parts. Each part represents the results of a dataset. The best performances are indicated by numbers in bold. We mean here by the best performance the highest precision, recall, F-measure and accuracy. This is followed by the number of selected genes. That means when the results obtained using different feature selection techniques equals, we consider the technique with smallest number of genes is the best. For example, in the case of Leukemia-subtype the results obtained using the ChiS equal to the results obtained using the combined technique (ChiSIG). However, we considered the results obtained using the combined technique are better because it outperformed the ChiS technique in the term of the number of selected genes. The number of genes in the case of ChiS is 300 while in the case of the combined technique is 233 genes. This is true regarding other cases.

For MLP, the best performance is obtained using the first structure, defined above, for all datasets. For SMO-SVM the best performance is obtained when using Poly Kernel function for Leukemia-subtype and Leukemia-ALLAML and PUK kernel function for colon dataset.

Table 4. The results obtained using MLP

	Leukemia-subtype				Leukemia-ALLAML				Colon			
	Prec	Recall	F	Acc	Prec	Recall	F	Acc	Prec	Recall	F	Acc
ChiS	96.8	96.4	96.5	96.43	94.7	94.1	94.0	94.12	84.1	84.2	83.9	84.21
ChiS-SMOTE	96.4	96.4	96.4	96.43	94.7	94.1	94.0	94.12	91.0	89.5	89.0	89.47
IG	94.8	94.6	94.7	94.64	94.7	94.1	94.0	94.12	84.2	78.9	76.2	78.95
IG-SMOTE	95.8	95.5	95.6	95.54	97.2	97.1	97.0	97.06	91.0	89.5	89.0	89.47
ChisIG	96.8	96.4	96.5	96.43	97.2	97.1	97.0	97.06	87.4	84.2	82.9	84.21
ChisIG-SMOTE	94.8	94.6	94.7	94.64	97.2	97.1	97.0	97.06	91.0	89.5	89.0	89.47

3.6 Analysis and Comparison

To analysis the effect of the SMOTE techniques, we classified its effects into three types: positive effect, negative effect and no effect. The experimental results shown in Tables 4, 5 and 6 demonstrate that 19 cases out of 27 cases

Table 5. The results obtained using SMO-SVM

	Leukemia-subtype				Leukemia-ALLAML				Colon			
	Prec	Recall	F	Acc	Prec	Recall	F	Acc	Prec	Recall	F	Acc
ChiS	96.3	95.5	95.7	95.54	88.2	85.3	84.5	85.29	87.4	84.2	82.9	84.21
ChiS- SMOTE	95.6	95.5	95.5	95.54	94.7	94.1	94.0	94.12	91.0	89.5	89.0	89.47
IG	95.9	95.5	95.6	95.54	97.2	97.1	97.0	97.06	81.4	73.7	68.6	73.68
IG-SMOTE	96.6	96.4	96.4	96.43	97.2	97.1	97.0	97.06	84.1	84.2	83.9	84.21
ChisIG	97.3	96.4	96.6	96.43	94.7	94.1	94.0	94.12	78.9	68.4	59.7	68.42
ChisIG-SMOTE	97.6	97.3	97.3	97.32	90.2	88.2	87.8	88.24	91.0	89.5	89	89.47

Table 6. The results obtained using Random forests

	Leukemia-subtype				Leukemia-ALLAML				Colon			
	Prec	Recall	F	Acc	Prec	Recall	F	Acc	Prec	Recall	F	Acc
ChiS	89.0	93.8	91.3	93.75	88.2	85.3	84.5	85.29	79.1	78.9	78.0	78.95
ChiS-SMOTE	96.6	96.4	96	96.43	92.3	91.2	91.0	91.18	91.0	89.5	89.0	89.47
IG	89.4	93.8	91.3	93.75	90.2	88.2	87.8	88.24	81.4	73.7	68.6	73.68
IG-SMOTE	95.1	94.6	93.9	94.64	88.2	85.3	84.5	85.29	87.4	84.2	82.9	84.21
ChisIG	90.0	94.6	92.2	94.64	86.4	82.4	81.1	82.35	79.1	78.9	78.0	78.95
ChisIG-SMOTE	94.9	95.5	94.9	95.54	90.2	88.2	87.8	88.24	87.4	84.2	82.9	84.21

SMOTE technique has positive effect in which the results were improved. Additionally, SMOTE technique has negative effect in three cases and has no effects in five cases. These cases are summarized in Table 7. In general, the overall average of results obtained using all classifiers is improved after applying SMOTE as shown in Figure 2.

Table 7. The effects of the SMOTE (+ means positive effect, - means negative and = means no change)

Classifier	Feature Selec- tion	Leukemia- subtype	Leukemia- ALLAML	Colon
MLP	ChiS	=	=	+
	IG	+	+	+
	ChiSIG	-	=	+
SMO	ChiS	=	+	+
	IG	+	=	+
	ChiSIG	+	-	+
RF	ChiS	+	+	+
	IG	+	-	+
	ChiSIG	+	+	+

The sensitivity of the classifiers to the SMOTE technique is analysed. The experimental results indicate that SMO-SVM and Random forests classifiers are more sensitive positively to SMOTE. Figure 3 shows that the overall average of

the improvements in the results (accuracy rates) obtained using SMO-SVM and Random forests are more significant than MLP.

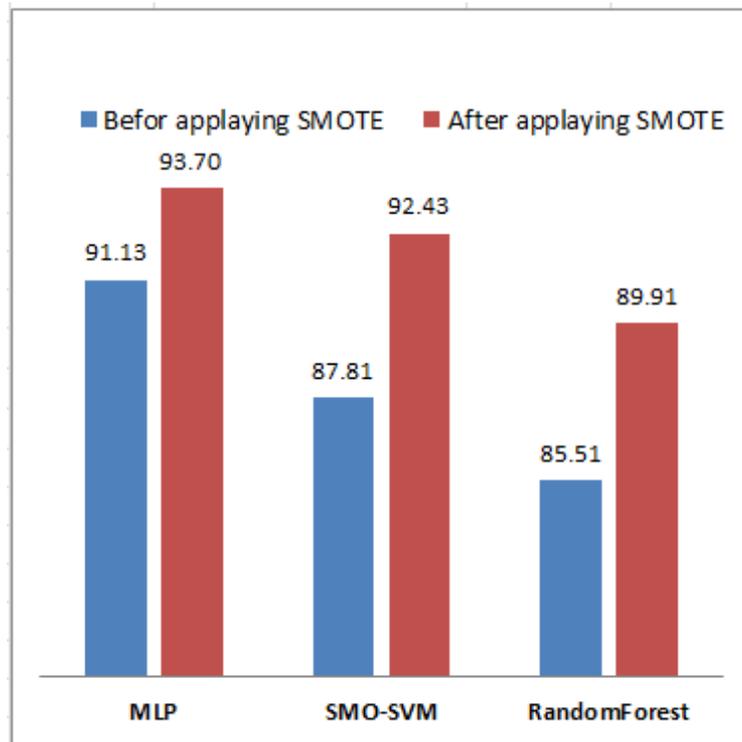


Fig. 2. The overall average accuracy of MLP, SMO-SVM and Random forests classifiers obtained in aforementioned experiments

One of the main objectives of this research is the focus on the ensemble of learning methods. Therefore, other experiments were conducted using Random forests and the ensemble of k-NN classifiers. ChiS feature selection is used with using 10-fold-cross validation mode (instead of training and testing mode) with applying SMOTE. The most informative genes for the three datasets were first investigated. The number of the most informative genes with the accuracies using Random forests classifier are shown in Fig. 3. The best performances of the most significant set of genes obtained using Random forests classifier is illustrated in Table 8. The aforementioned procedure is also followed using the ensemble of k-NN classifiers instead of Random forests. The number of the most informative genes with the accuracies using the ensemble of k-NN classifiers are shown in Fig. 4. Table 9 illustrates the best performances of the most significant set of genes obtained using the ensemble of k-NN classifiers.

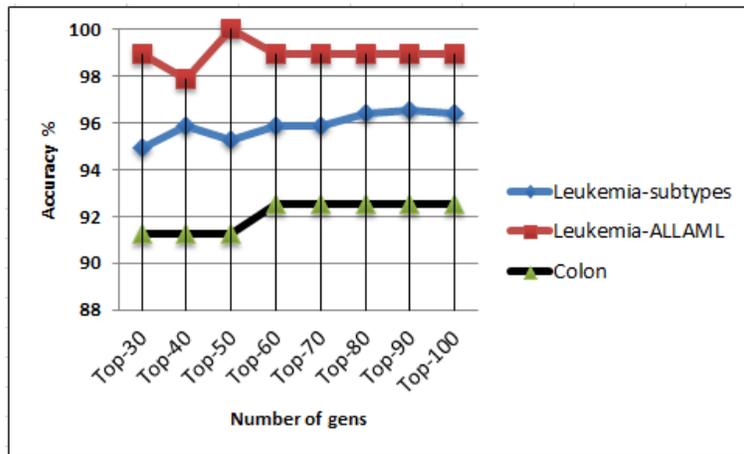


Fig. 3. Accuracy of Random forests classifier with ChiS feature selection on the three datasets handled using SMOTE

Table 8. The most significant set of genes selected by ChiS feature selection after applying SMOTE using Random forests

Dataset	#genes	Prec	Recall	F	Acc
Leukemia-subtypes	90	96.5	96.6	96.5	96.6
Leukemia-ALLAML	50	100.0	100.0	100.0	100.0
Colon	60	92.5	92.5	92.5	92.5

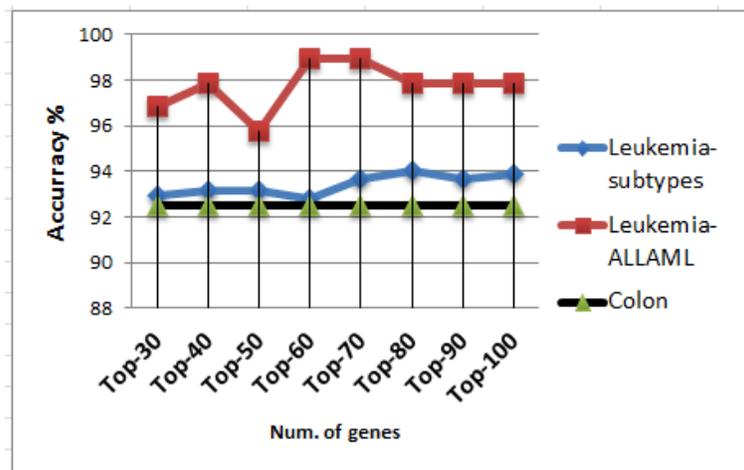


Fig. 4. Accuracy of the ensemble of k-NN classifiers with ChiS feature selection on the three datasets handled using SMOTE

Table 9. The most significant set of genes selected by ChiS feature selection after applying SMOTE and evaluated using the ensemble of k-NN classifiers

Dataset	#genes	Prec	Recall	F	Acc
Leukemia-subtypes	100	94.2	93.9	93.5	93.85
Leukemia-ALLAML	60	99.0	98.9	98.9	98.94
Colon	30	93.1	92.9	92.5	92.95

We compared the obtained results with related works based on some criteria as illustrated in Table 10.

Table 10. Comparisons of our work and related works

Ref	Techniques	#genes	Experimental method	Acc%
Dataset: Leukemia-subtype				
Li Zhang [25]	similarity-balanced discriminant neighborhood embedding (SBDNE)		Holdout method	86.95
Our work	ChisIG-SMOTE and SMO	255	Holdout method	97.32
Our work	ChisIG-SMOTE and SMO	90	10-fold-CV	96.56
Dataset: Leukemia-ALLAML				
Li Zhang [25]	Locally linear discriminant embedding (LLDE)		Holdout method	88.18
Lee et al. [6]	Neural network-based finite impulse response extreme learning machine (FIR-ELM)		10-fold-CV	96.53
Our work	(1) ChiSIG and MLP (2) ChiSIG-SMOTE and MLP	145	Holdout method	97.06
Our work	ChiSIG-SMOTE Random forests	50	10-fold-CV	99.95
Dataset: Colon				
Rathore et al. [8]	Sequence of Chi-S and mRMR) & SVM	40	10-fold-CV	91.94
Lotfi and Ke-shavarz [7]	PCA-BEL		10-fold-CV	87.40
Lee et al. [6]	FIR-ELM		10-fold-CV	79.76
Our work	(1) ChisIG-SMOTE and MLP (2) ChisIG-SMOTE and SMO	111	Holdout method	89.47
Our work	ChiS-SMOTE and ensemble of k-NN classifiers	30	10-fold-CV	92.50

4 Conclusion

A research is conducted to classify cancers using gene expression datasets. Two main issues are addressed: the class imbalance and the curse of dimensionality of gene expression microarray datasets. We addressed the class imbalance problem through over-sampling the minority classes using SMOTE technique. While the issue of the curse of dimensionality was addressed in this research by applying two popular feature selection techniques individually and in combination. Three publicly available bio-medical datasets (Leukemia-subtype, Leukemia-ALLAML and Colon tumor) which suffer the problems of the class imbalance and curse of dimensionality were used. The effectiveness of the proposed approach was evaluated using MLP, SMO-SVM and Random forests classification algorithms. A majority voting based ensemble of k-NN classifiers is also proposed and evaluated. The experimental results indicate that SMOTE technique results in improving the results in the most cases. Additionally the highest results are obtained using the combined feature selection technique in the most cases. Moreover, the sensitivity of the used classifiers to the SMOTE is also analysed in this research. The experimental results demonstrates that SMO-SVM and Random forests are more sensitive positively to the effect of SMOTE than MLP. For Leukemia-subtype, the best performance is obtained using SMO-SVM with the combined feature selection technique and after applying SMOTE such that an accuracy rate of 97.32% is obtained. For the second dataset Leukemia-ALLAML, the highest accuracy of 97.06% is obtained using MLP with the combined feature selection technique and SMO-SVM with IG feature selection technique. For the colon dataset, the highest accuracy rate of 89.47% is achieved using both MLP and SMO-SVM with the combined feature selection technique after applying SMOTE. The same results were obtained using Random forests using the ChiS after applying SMOTE. Other experiments were performed to investigate the most significant set of genes for all datasets using ChiS feature selection and ensemble learning methods (Random forests and the proposed ensemble of k-NN classifiers) after applying SMOTE. The highest accuracy rates of Leukemia-subtype was 96.56% which is obtained using a set of 90 genes, 100% which is obtained using a set of 50 genes for Leukemia-ALLAML and the Random forests and 92.50% which is obtained using a set of 30 genes for the colon dataset and the proposed ensemble of k-NN classifiers. The obtained results are compare favourably with the literature. The authors are working now on investigating other techniques to combine feature selection techniques and proposing a heterogeneous ensemble learning methods based on meta learner.

References

1. Sridhar Ramaswamy, Pablo Tamayo, Ryan Rifkin, Sayan Mukherjee, Chen-Hsiang Yeang, Michael Angelo, Christine Ladd, Michael Reich, Eva Latulippe, Jill P Mesirov, et al. Multiclass cancer diagnosis using tumor gene expression signatures. *Proceedings of the National Academy of Sciences*, 98(26):15149–15154, 2001.
2. Robert Tibshirani, Trevor Hastie, Balasubramanian Narasimhan, and Gilbert Chu. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proceedings of the National Academy of Sciences*, 99(10):6567–6572, 2002.
3. Todd R Golub, Donna K Slonim, Pablo Tamayo, Christine Huard, Michelle Gaasenbeek, Jill P Mesirov, Hilary Coller, Mignon L Loh, James R Downing, Mark A Caligiuri, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *science*, 286(5439):531–537, 1999.
4. Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, pages 321–357, 2002.
5. Wei-Jiun Lin and James J Chen. Class-imbalanced classifiers for high-dimensional data. *Briefings in bioinformatics*, page bbs006, 2012.
6. Kevin Lee, Zhihong Man, Dianhui Wang, and Zhenwei Cao. Classification of bioinformatics dataset using finite impulse response extreme learning machine for cancer diagnosis. *Neural Computing and Applications*, 22(3-4):457–468, 2013.
7. Ehsan Lotfi and Azita Keshavarz. Gene expression microarray classification using pca-bel. *Computers in biology and medicine*, 54:180–187, 2014.
8. Saima Rathore, Muhammad Aksam Iftikhar, and Mutawarra Hussain. A novel approach for automatic gene selection and classification of gene based colon cancer datasets. In *Emerging Technologies (ICET), 2014 International Conference on*, pages 42–47. IEEE, 2014.
9. Saima Rathore, Mutawarra Hussain, and Asifullah Khan. Gecc: gene expression based ensemble classification of colon samples. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, 11(6):1131–1145, 2014.
10. Sara Haddou Bouazza, Nezha Hamdi, Abdelouhab Zeroual, and Khalid Auhmani. Gene-expression-based cancer classification through feature selection with knn and svm classifiers. In *Intelligent Systems and Computer Vision (ISCV), 2015*, pages 1–6. IEEE, 2015.
11. Haider Banka and Suresh Dara. A hamming distance based binary particle swarm optimization (hdbpso) algorithm for high dimensional feature selection, classification and validation. *Pattern Recognition Letters*, 52:94–100, 2015.
12. Monika Simjanoska, Ana Madevska Bogdanova, and Zaneta Popeska. Recognition of colorectal carcinogenic tissue with gene expression analysis using bayesian probability. *ICT Innovations*, pages 305–314, 2012.
13. Ana Madevska Bogdanova, Monika Simjanoska, and Zaneta Popeska. Classification of colorectal carcinogenic tissue with different dna chip technologies. In *the 6th International Conference on Information Technology, ser. ICIT*, 2013.
14. Wing-Cheong Wong, Marie Loh, and Frank Eisenhaber. On the necessity of different statistical treatment for illumina beadchip and affymetrix genechip data and its significance for biological interpretation. *Biology direct*, 3(1):23, 2008.
15. Monika Simjanoska, Ana Madevska Bogdanova, and Zaneta Popeska. Bayesian posterior probability classification of colorectal cancer probed with affymetrix microarray technology. In *Information & Communication Technology Electronics & Microelectronics (MIPRO), 2013 36th International Convention on*, pages 959–964. IEEE, 2013.

16. Monika Simjanoska and Ana Madevska Bogdanova. Novel methodology for crc biomarkers detection with leave-one-out bayesian classification. In *ICT Innovations 2014*, pages 225–236. Springer, 2015.
17. Muchenxuan Tong, Kun-Hong Liu, Chungui Xu, and Wenbin Ju. An ensemble of svm classifiers based on gene pairs. *Computers in biology and medicine*, 43(6):729–737, 2013.
18. Jin Cao, Li Zhang, Bangjun Wang, Fanzhang Li, and Jiwen Yang. A fast gene selection method for multi-cancer classification using multiple support vector data description. *Journal of biomedical informatics*, 53:381–389, 2015.
19. Eng-Juh Yeoh, Mary E Ross, Sheila A Shurtleff, W Kent Williams, Divyen Patel, Rami Mahfouz, Fred G Behm, Susana C Raimondi, Mary V Relling, Anami Patel, et al. Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. *Cancer cell*, 1(2):133–143, 2002.
20. Uri Alon, Naama Barkai, Daniel A Notterman, Kurt Gish, Suzanne Ybarra, Daniel Mack, and Arnold J Levine. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proceedings of the National Academy of Sciences*, 96(12):6745–6750, 1999.
21. Haibo He, Edwardo Garcia, et al. Learning from imbalanced data. *Knowledge and Data Engineering, IEEE Transactions on*, 21(9):1263–1284, 2009.
22. Ian H Witten and Eibe Frank. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2005.
23. John Platt et al. Fast training of support vector machines using sequential minimal optimization. *Advances in kernel methodssupport vector learning*, 3, 1999.
24. Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.
25. Li Zhang, Liqiang Qian, Chuntao Ding, Weida Zhou, and Fanzhang Li. Similarity-balanced discriminant neighbor embedding and its application to cancer classification based on gene expression data. *Computers in biology and medicine*, 64:236–245, 2015.

Evaluation of differential gene expression workflow variability through RNA-Seq experiments.

Alvaro Pérez-Sala¹ and Ignacio M. Larráyo¹

¹ Centre for Biomedical Research of La Rioja (CIBIR), Fundación Riojasalud, Logroño, La Rioja, 26006, Spain
(aperez, ilarrayoz)@riojasalud.es

Keywords: Bioinformatics, Sequence analysis, RNA-Seq.

1 Introduction

Recently the number of researchers using high throughput sequencing techniques in their studies has grown significantly, and high throughput RNA-Seq has validated its utility across a wide set of experiments. Moreover, differential gene expression analysis has been one of the most important procedures to analyse mRNA expression level between different conditions.

RNA-Seq data analysis require specific bioinformatic processes depending on the experimental paradigm. This study focuses on the bioinformatic protocols for differential gene expression quantification between different conditions. Differential expression analysis workflows have evolved including new features, that aim to improve the specificity and sensibility of these analyses. Also, these workflows not only improved by giving more accurate results, but also have incorporated new algorithms reducing computational requirements. The analysis of strengths and weakness between different protocols could be interesting for the scientific community.

2 Material and methods

Already described protocols like “tuxedo” (1)(TopHat, Cufflinks and CummeRbund) and “new tuxedo” (2)(HISAT, StringTie and Ballgown) have been chosen in addition to protocols composed by general purpose programs (3) (STAR (4), FeatureCounts (5), and EdgeR (6) / DeSeq (7)) for this study. FeatureCounts, DeSeq and EdgeR have been combined with data generated by all aligners to study variations on the results through the different tools.

Many features, such as memory effectivity, performance as mapping speed, etc. of these three different analysis workflows has been evaluated across the three main protocol steps looking for differences, advantages and disadvantages of each one. Meanwhile we have cross combined the three main protocol steps between them to compare the combined performance versus the original workflows. For final workflow evaluation significantly differential expressed gene sets identified for each method, were studied in GeneMania (8) analysing predicted pathways by the three different gene sets.

2

3 Results and discussion

After aligners evaluation, STAR got the best average read mapping ratio (99%) compared with TopHat (94,5%) and HiSat (96,5%). Other parameters like CPU and memory requirements have been measured in addition of mapping read ratio. FeatureCounts was the quickest in the second step and provides a simple and easy to understand dataset. Finally, EdgeR, using FeatureCounts, presents advantages in flexibility at the experimental design, reliability of statistical models using empiric bayes methods, and normalization methods (TMM, Trimmed mean of m-values).

To sum up, we propose STAR, FeatureCounts and EdgeR workflow as a low computational requirement and reliable protocol in differential gene expression analysis with high experimental design flexibility.

4 Thanks

This work was supported by Instituto de Salud Carlos III CP15/00198, Fondo Social Europeo (Periodo de Programación 2014 – 2020) and Fondo Europeo de Desarrollo Regional granted to IML.

References

1. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* [Internet]. 2012;7(3):562–78. Available from: <http://dx.doi.org/10.1038/nprot.2012.016><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3334321&tool=pmcentrez&rendertype=abstract>
2. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc* [Internet]. 2016;11(9):1650–67. Available from: <http://dx.doi.org/10.1038/nprot.2016.095><http://10.1038/nprot.2016.095><http://www.nature.com/nprot/journal/v11/n9/abs/nprot.2016.095.html#supplementary-information>
3. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* [Internet]. 2014;15(12):550. Available from: <http://genomebiology.biomedcentral.com/articles/10.1186/s13059-014-0550-8>
4. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* [Internet]. 2013 Jan;29(1):15–21. Available from: <https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/bts635>
5. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* [Internet]. 2014 Apr 1;30(7):923–30. Available from: <https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btt656>

6. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* [Internet]. 2010 Jan 1;26(1):139–40. Available from: <https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btp616>
7. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol* [Internet]. 2010;11(10):R106. Available from: <http://genomebiology.com/2010/11/10/R106>
8. Montojo J, Zuberi K, Rodriguez H, Kazi F, Wright G, Donaldson SL, et al. GeneMANIA cytoscape plugin: Fast gene function predictions on the desktop. *Bioinformatics*. 2010;26(22):2927–8.

New features and recipes for simulating structural mutations with BAMSurgeon

Adam D. Ewing

Mater Research Institute - University of Queensland
TRI Building Level 4, Woolloongabba, QLD, Australia
adam.ewing@mater.uq.edu.au

1 Introduction

The computational methods used to detect mutations from next generation sequence data are always improving as is the throughput and accuracy of the underlying data. As technology advances so does our ability ask more nuanced and interesting questions. Examples include detection of somatic mutations at low variant allele frequency (VAF), detection of mutations in single-cell sequence data, and the ability to resolve complex structural rearrangements. With increasing sophistication and difficulty in detecting mutations comes increasing difficulty in validating mutations and in tuning software parameters. In many cases false positives can be detected through follow-up with molecular methods (PCR, ddPCR), but tuning parameters to control the false negative rate is significantly more challenging.

One approach to this challenge is realistic simulation of mutations using BAMSurgeon, a tool which adds mutations to existing data. Using existing data has the advantage of side-stepping the need to simulate the underlying error rate associated with sequencing technology, and also allows direct comparison of sequencing methods. BAMSurgeon is highly configurable allowing the user to specify mutation locations, types, VAFs, and meta-information specific to various mutation types. While the basic functionality of BAMSurgeon has been described elsewhere [1,2], a number of recently developed novel features unique to BAMSurgeon which enable simulation of a wider repertoire of structural variation are described here.

2 Accurate Simulation of Retrotransposon Insertions

Retrotransposons are among the most potent mutagens in healthy mammalian genomes and in many tumour types, accounting for the large majority of structural insertion mutations when indels are excluded [3]. Detecting these mutations is challenging due to their repetitive nature - many copies already exist in the genome - and the specifics of their mechanism of mobilisation also present unique challenges [4]. From a sequence analysis standpoint, canonical retrotransposon insertions consist of two breakpoints at either end of the element and the

insertion is typically flanked by target-site duplications (TSD). The 3' end of the inserted sequence is typically polyadenylated and the 5' end may be truncated or inverted relative to the donor sequence. Typically the insertions occur preferentially at endonuclease cleavage sites, and the sequence and specificity are dependent on the retroelement family. BAMSurgeon includes the necessary functionality to accurately simulate retrotransposon insertions including target site selection, TSD generation, and scripts to assist in generating custom libraries of insertion sequences with realistic characteristics including truncations, 5' inversions, and polyadenylation.

3 Simulation of Complex Rearrangements

BAMSurgeon supports a number of simple structural mutation types: duplication, deletion, insertion, inversion, and translocation [2]. It also allows for the chaining of these mutations together to create complex structural rearrangements e.g. invert part of a duplicated sequence and create a deletion at one of the breakpoints. This functionality can be used to simulate some of the complex rearrangements typical in hypermutated tumour genomes including chromothripsis, which is thought to be associated with DNA repair defects, and chromoplexy, which is associated with mitotic defects. Chromothripsis is observed as a large number of proximal rearrangements, while chromoplexy involves a large number of distal and inter-chromosomal rearrangements. BAMSurgeon, in conjunction with included auxiliary scripts, includes the functionality to generate examples of chromothripsis and chromoplexy and simulated them in existing genomes, providing a much-needed means to test the accuracy of tools in deciphering these complex mutations.

References

1. Ewing AD, Houlahan KE, Hu Y, Ellrott K, Caloian C, Yamaguchi TN, et al.: Combining tumor genome simulation with crowdsourcing to benchmark somatic single-nucleotide-variant detection. *Nat. Methods.* 12, 623–630 (2015)
2. Lee AY-W, Ewing AD, Ellrott K, Hu Y, Houlahan KE, Christopher Bare J, et al.: Combining accurate tumour genome simulation with crowd sourcing to benchmark somatic structural variant detection. *bioRxiv.* p. 224733. doi:10.1101/224733 (2017)
3. Rodriguez-Martin B, Alvarez EG, Baez-Ortega A, Demeulemeester J, Ju YS, Zamora J, et al.: Pan-cancer analysis of whole genomes reveals driver rearrangements promoted by LINE-1 retrotransposition in human tumours. *bioRxiv.* p. 179705. doi:10.1101/179705 (2017)
4. Ewing AD: Transposable element detection from whole genome sequence data. *Mobile DNA.* 6, 24 (2015)

The roles and mechanisms of celecoxib in gastric cancer
NCI-N87 cells

Abstract

Background

Gastric cancer is a common malignancy worldwide, characterized by highly invasiveness and aggressiveness. Besides operation, chemotherapy is usually used to perform adjuvant therapy for gastric cancer. Chemotherapy drugs can be divided into cytotoxic drugs and non cytotoxic drugs, and the curative effect of non cytotoxic drugs for the treatment of gastric cancer is superior to cytotoxic drugs with lower side effect. Non-steroidal anti-inflammatory drugs including cyclooxygenase-2 inhibitors belong to non cytotoxic drugs and can promote the treatment of many tumors. Celecoxib as a kind of cyclooxygenase-2 inhibitor may inhibit proliferation and angiogenesis, as well as promote apoptosis in the tumor cells. However, the celecoxib for inhibiting the occurrence and development of gastric cancer is still unclear.

Objective

1. The effect of celecoxib on cell viability, cell cycle and cell apoptosis on human gastric carcinoma cell lines NCI-N87 were investigated.

2. We verified the mRNA levels of COX-2 in NCI-N87 cells treated

and untreated with celecoxib.

3. The RNA sequencing data of NCI-N87 human gastric carcinoma cells treated with or without celecoxib were prepared. Then, differentially expressed genes (DEGs) and lncRNAs were identified for pathway enrichment analysis. Afterward, protein-protein interaction (PPI) network for DEGs was constructed and module analysis was performed. Additionally, co-expression analysis of differentially expressed genes and lncRNAs was performed. This study is aimed to investigate the mechanisms of lncRNAs in gastric cancer cell line treated with celecoxib.

4. We verified the mRNA levels of lnc-SCD-1:13 and lnc-PTMS-1:3 as well as ITGA3 and DVL1, aiming to further reveal whether lncRNAs involved in the treatment of celecoxib in NCI-N87 cells.

Methods

1. MTT assay was used to detect cell viability of NCI-N87 cells. Besides, cell cycle and apoptosis were observed by flow cytometry analysis.

2. The mRNA levels of COX-2 were detected by Real time-PCR in NCI-N87 cells treated and untreated with celecoxib.

3. RNA libraries for NCI-N87 cells treated and untreated with celecoxib were built and the next generation sequencing was performed. Quality control (QC) of obtained next generation sequencing (NGS) data was conducted by NGS QC Toolkit. Cuffdiff was applied to screen

differentially expressed genes and lncRNAs with the cut-off criteria of $q\text{-value} < 0.05$. Gene Ontology (GO) functional and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis were performed using GO-function package and KEGGprofile in Bioconductor, respectively ($p < 0.05$ and gene counts ≥ 2). The PPIs for DEGs were predicted using version 9.1 of STRING database with the combine score > 0.7 . Pearson correlation coefficients between DEGs and lncRNAs were firstly calculated. The co-expressed genes and lncRNAs pairs were selected with $|\text{Pearson correlation coefficient}| > 0.98$.

4. With key genes in NCI-N87 cells treated with celecoxib were selected as experimental objects, the mRNA levels of lnc-SCD-1:13 and lnc-PTMS-1:3 as well as ITGA3 and DVL1 were detected by Real time-PCR in NCI-N87 cells.

Results

1. Effect of celecoxib on cell proliferation, cell cycle and cell apoptosis in NCI-N87 cells

The results of MTT analysis showed that the cell viability was significantly inhibited in celecoxib treated cells group compared with untreated cells ($p < 0.05$). The results of flow cytometry suggested that compared with control group, celecoxib treated cells had a significantly increased percentage of cells in G0/G1 phase ($p < 0.05$) and lower percentage of cells in S and G2 phase ($p < 0.05$). The results of flow

cytometry showed that the cell viability in early and late apoptosis were both significantly increased in celecoxib treated cells compared with untreated cells ($p < 0.05$).

2. Effect of celecoxib on the mRNA levels of COX-2 in NCI-N87 cells

We found that the expression of COX-2 mRNA in celecoxib with effective dose treated cells had no significant difference compared with untreated cells.

3. Differentially expressed genes and lncRNAs

A total of 490 DEGs, including 302 up-regulated and 188 down-regulated DEGs, were identified between the celecoxib group and the control group. Meanwhile, 37 differentially expressed lncRNAs, including 19 up-regulated and 18 down-regulated differentially expressed lncRNAs, were screened.

4. Functional enrichment analysis for DEGs

GO enrichment analysis revealed that the up-regulated DEGs mainly were enriched in small molecule metabolic process, involving in ABCC3, ACAA1, B3GNT3 and CD320, and extracellular region, related to ITGA3, ITGA6, ITGB4 and ITGB5 respectively. The down-regulated DEGs were enriched in tissue development (ADAM9, ALDH1A3 and FNDC3B), extracellular region (ADAM9, CCDC80 and CLIC5) and chemokine activity (ADAM9, HSPB1, MARCKS and PKP2),

respectively.

5. Pathway enrichment analysis for DEGs

Pathway enrichment analysis indicated that the up-regulated DEGs were significantly enriched in the metabolic pathways (ATP5G1, ATP5G3, ATP6AP1, COX8A, CYC1, NDUFS3, UQCRC1, UQCRC2 and UQCRFS1) and oxidative phosphorylation (ATP5G1, ATP5G3, ATP6AP1, COX8A, CYC1, NDUFS3, UQCRC1, UQCRC2 and UQCRFS1). On the other hand, the down-regulated DEGs were enriched in epithelial cell signaling in helicobacter pylori infection (CXCL1, CXCL8 and MAP3K14), chemokine signaling pathway (BCAR1, CXCL1, CXCL3, CXCL5 and CXCL8) and cytokine-cytokine receptor interaction (BMPR2, CXCL1, CXCL3, CXCL5, CXCL8 and TNFRSF19).

6. PPI network and module analysis

After the PPIs of DEGs was predicted, two modules (Module 1 and Module 2) with the highest significance were selected. The DEGs in module 1 (ITGB6, LAMA3, ITGA6, ITGB4, ITGB5, ITGA3 and ITGB8) were mainly related to functions about integrin-mediated signaling pathway, extracellular matrix organization and cell adhesion. In module 2, DEGs, including CYC1, COX8A, UQCRC1, NDUFS3, UQCRC2 and UQCRFS1, were involved in the respiratory electron transport chain and mitochondrial inner membrane. The DEGs (such as LAMA3, ITGA3,

ITGA6, ITGB4, ITGB5, ITGB6 and ITGB8) in the module 1 were mainly enriched in the focal adhesion pathway and ECM-receptor interaction pathway. Meanwhile, three up-regulated DEGs (LAMA3, ITGA6 and ITGA3) were enriched in pathways in cancer. The seven up-regulated DEGs including ATP5G3, CYC1, COX8A, UQCRC1, NDUFS3, UQCRC2 and UQCRFS1 in module 2 were enriched in oxidative phosphorylation and metabolic pathways.

7. Co-expression analysis of differentially expressed genes and lncRNAs

The DEGs co-expressed with lnc-SCD-1:13, lnc-LRR1-1:2, lnc-PTMS-1:3, lnc-S100P-3:1, lnc-AP000974.1-1:1 and lnc-RAB3IL1-2:1 were enriched in the pathways related to cancer, such as basal cell carcinoma (FZD7, DVL1 and WNT7B), pathways in cancer (FZD7, DVL1 and ITGA3) and ECM-receptor interaction (ITGA3). The DEGs (DVL1, NFAT5, WNT11 and WNT7B) co-expressed with lnc-SCD-1:13, lnc-LRR1-1:2 and lnc-S100P-3:1 were enriched in Wnt signaling pathway. The DEGs (BMP4, FZD7, WNT11 and WNT7B) co-expressed with lnc-SCD-1:13, lnc-LRR1-1:2, lnc-PTMS-1:3, lnc-S100P-3:1 and lnc-AP000974.1-1:1 were enriched in Hedgehog signaling pathway.

8. Effect of celecoxib on the mRNA levels of lnc-SCD-1:13 and lnc-PTMS-1:3 as well as ITGA3 and DVL1 in NCI-N87 cells

RT-PCR analysis demonstrated increased mRNA levels of these

lnc-RNAs and genes in celecoxib treated cells compared with untreated cells ($p < 0.05$).

Conclusion

1. Low concentration of celecoxib plays an important role of anti-cancer via regulating the cell cycle and apoptosis for gastric cancer by mediating the other mechanisms rather than the dependency mechanism of COX-2.

2. Celecoxib may reduce the cell migration and proliferative activity to inhibit the development of gastric cancer by downregulation of CXCL1, CXCL3, CXCL5 and CXCL8, which are participated in the chemokine signaling pathway and cytokine-cytokine receptor interaction.

3. Celecoxib can reduce the expression of WNT7B, and may inhibit the progression of gastric cancer via suppressing Wnt signaling pathway.

4. Celecoxib can inhibit the development of gastric cancer by ATP5G1, ATP5G3, COX8A, CYC1, NDUFS3, UQCRC1, UQCRC2 and UQCRFS1, which are participated in the oxidative phosphorylation.

5. The DEGs co-expressed with lnc-SCD-1:13, lnc-LRR1-1:2, lnc-PTMS-1:3, lnc-S100P-3:1, lnc-AP000974.1-1:1 and lnc-RAB3IL1-2:1 may play a key role on the development of gastric cancer via participating in many pathways.

6. ITGA3 as co-expression gene with lnc-SCD-1:13 and lnc-PTMS-1:3 may play a key role on the development of gastric cancer

via participating in the integrin-mediated signaling pathway.

Keywords:

Gastric cancer, Celecoxib, Differentially expressed genes, Long non-coding RNAs, Cyclooxygenase-2, Anti-cancer effect