IYBSSD2022

XXVIII Symposium on Bioinformatics and Computer-Aided Drug Discovery

PROCEEDINGS BOOK

Institute of Biomedical Chemistry
Moscow, Russia (Virtual), May 24-26, 2022

Russian Academy of Sciences
Ministry of Science and Higher Education of Russian Federation
Institute of Biomedical Chemistry
Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry
of the Russian Academy of Sciences

Committee

Organization committee:

Vladimir Poroikov (Chairman)
Institute of Biomedical Chemistry, Russia

Roman Efremov (Chairman)
Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry of the Russian Academy of Sciences, Russia

Scientific committee:

James Devillers
CTIS, Rillieux La Pape, France

Frank Eisenhaber
A*STAR Bioinformatics Institute, Singapore

Rajesh Kumar Goel
Punjabi University - Patiala, India

Athina Geronikaki
Aristotle University of Thessaloniki, Greece

Alexey Lagunin
Pirogov Russian National Research Medical University, Russia

Uko Maran
University of Tartu, Estonia

José L. Medina-Franco
National Autonomous University of Mexico, Mexico

Kyoung Tai No
Yonsei Univeristy, Republic of Korea

Tudor Oprea
University of New Mexico, USA

Dmitry Osolodkin
Chumakov Federal Scientific Center for Research and Development of Immune-and-Biological Products of Russian Academy of Sciences

Vladimir Palyulin
Faculty of Chemistry - Lomonosov Moscow State University, Russia

Pavlo Polishchuk
Palacky University and University Hospital in Olomouc, Czech Republic

Kunal Roy
Jadavpur University, India

Garikapati Narahari Sastry
CSIR-North East Institute of Science and Technology, India

Marcus Scotti
Universidade Federal da Paraíba, Brasil

Hanoch Senderowitz
Bar-Ilan University, Israel

Igor Tetko
Institute of Structural Biology - Helmholtz Zentrum München, Germany

Vyacheslav Tribulovich
Saint-Petersburg State Institute of Technology, Russia

Anna Tsantili-Kakoulidou
National & Kapodistrian University of Athens, Greece

Pavel Vassiliev
Volgograd State Medical University, Russia

Weiliang Zhu
Shanghai Institute of Materia Medica, Chinese Academy of Sciences, China

Local Organizing Committee:

Tatyana Pleshakova (Chair)
Institute of Biomedical Chemistry

Polina Savosina (Secretary)
Institute of Biomedical Chemistry

Dmitry Druzhilovskiy
Institute of Biomedical Chemistry

Anastasia Rudik
Institute of Biomedical Chemistry

Vladlen Skvortsov
Institute of Biomedical Chemistry

Kira Stefanovich
Institute of Biomedical Chemistry

DOI: 10.18097/BCADD2022

ISBN 978-5-900760-18-6

PROCEEDINGS BOOK OF THE XXVIII SYMPOSIUM "BIOINFORMATICS AND COMPUTER-AIDED DRUG DISCOVERY" – Moscow: Institute of Biomedical Chemistry, 2022.

The materials of the XXVIII Symposium "Bioinformatics and Computer-Aided Drug Discovery" are presented. This Symposium is dedicated to the emerging challenges and opportunities for *in silico* drug discovery.

The Symposium's main topics: development and practical application of computational methods for finding and validation of new pharmacological targets, *in silico* design of potent and safe pharmaceutical agents, optimization of the structure and properties of drug-like compounds, rational approaches to the utilization of pharmacotherapeutic remedies in medical practice.

This information will be useful for researchers whose investigations are dedicated to creating computational methods and their application to drug research and development using bio- and chemoinformatics methods based on post-genomic technologies. It can also be useful for undergraduate, graduate, and postgraduate students specializing in the relevant fields.

Responsible editors: Corr. Member of Rus. Acad. Sci. V.V. Poroikov, Prof. R.G. Efremov

Dear Colleagues!

On behalf of the Organizing Committee and the Administration of the Institute of Biomedical Chemistry (IBMC), we are glad to welcome the participants of the XXVIII Symposium "Bioinformatics and Computer-Aided Drug Discovery".

For the first time, such a Symposium was held in 1995 in the framework of the II Russian National Congress "Man and Drugs" initiative by Full Member of the Russian Academy of Sciences (RAS) Alexander Archakov. Since then, this meeting has been held annually under the chairmanship of Corresponding Member of RAS Vladimir Poroikov - Head of the Bioinformatics Department at IBMC together with Full Member of RAS Nikolay Zefirov, and since 2018 - with Professor Roman Efremov. The annual Symposium provides an opportunity for researchers whose studies are related to the development and application of bioinformatics, chemoinformatics and computer-aided drug design methods, to exchange information on the latest achievements and discuss the modern trends for the development of this multidisciplinary field of science in the future.

Traditionally, a significant part of the scientific sessions of the Symposium, including the Young Scientists Contest, was held in the IBMC. In 2021, given the situation with the COVID-19 pandemic, the XXVII Symposium was held online, which allowed involvement of authoritative scientists from Germany, France and the USA. Taking into account this experience, it was decided to perform the XXVIII Symposium as virtual meeting. More than 400 staff scientists, PhD, graduate and undergraduate students from 49 countries have already registered to take part in the Symposium's sessions.

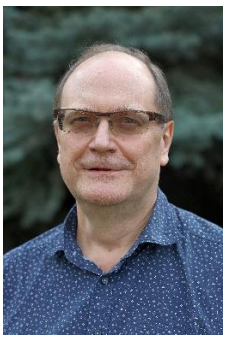
The main topics of the Symposium are especially important due to the active involvement of IBMC in the creation and development of the World-Class Scientific Center (NCMU) "Digital Biodesign and Personalized Health Care" within the framework of the Russian National Project "Science". This is a global project on the digitalization of healthcare and health management. Within the framework of this project, the development of anti-cancer drugs with the use of scaffoldless tissue engineered constructions and cellular spheroids for neogenesis is also carried out. As part of this project, IBMC will develop a digital information platform designed to optimize the decision-making by medical professionals regarding strategy and tactics of treatment using modern pharmacotherapy and taking into account the individual characteristics of the patient.

In 2022, the life of most of us has changed significantly: the people have lost the confidence in the future, and the usual work environment. This has led to an escalation of tenseness, and sometimes even to a complete rupture in relations between colleagues and within workgroups. Now, we are set at a challenge: how to hold out against the unfavorable conditions in order to continue our occupation with what we love, and to maintain high quality of our research? In this regard, the organization and holding this International Symposium is an important communicative step, which will allow the scientists to unite, helping them hear and understand each other.

I would like to thank the participants of the Symposium, wishing you a fruitful work!

Director of the Institute of Biomedical Chemistry
Doctor of Biological Sciences

Elena Ponomarenko



Dear Colleagues!

We are pleased to welcome you as participants of the XXVIII Symposium "Bioinformatics and Computer-Aided Drug Discovery".

Now all of us live and work in challenging conditions caused by coronavirus pandemics and global changes in the world that have been taking place recently. It has become evident that the emerging global threats to human health will continue to persist in the future. One of the critical issues in this situation is the rapid development and introduction into the medical practice of new - often innovative - medicines for the successful fight against newly emerging pathologies and known diseases, which manifestation is significantly changed due to these "newcomers". Solving this enormous task requires the mobilization of existing intellectual and material resources and the creation of new approaches to the discovery and development of new medicines.

As the practice of combating socially significant diseases in recent years, and especially COVID-19, shows, the use of multidisciplinary approaches of bioinformatics and computer-aided technologies in general plays a critical role in developing physiologically active compounds. These methods allow *in silico* processing big biomedical data, identifying promising molecular targets, elaborating quantitative structure-activity relationships, building and validating predictive models, and implementing lead compounds' rational molecular design. Altogether, this provides valid recommendations to medicinal chemists, pharmacologists, toxicologists, physicians, and regulatory authorities to optimize drug discovery, development, approval, and clinical usage.

Thus, our Symposium's topics and all of us are at the forefront of the ongoing protracted fight against global scale pathologies.

Let us use the Symposium discussion platform to exchange original scientific ideas, attractive methodological solutions, and breakthrough multidisciplinary technologies. This is especially important in connection with the recent world events, which complicate international scientific and educational relationships, efficient exchange of information and data. These factors have always been at the heart of scientific creativity, especially in the field of biomedicine. We believe that holding our Symposium in these difficult conditions, involving the participation of scientists from many countries, will help develop scientific diplomacy, preserve and increase professional and human relations of colleagues, establish new creative connections, and, as a result, increase the efficiency of computer technologies for the discovery of new medicines. We hope that our Symposium will also contribute to reducing tension in the world. The online format provides unique opportunities for this, including talks given by our authoritative colleagues from all over the world.

Welcome to the sessions of the XXVIII Symposium "Bioinformatics and Computer-Aided Drug Discovery". We wish you very exciting and fruitful meetings and discussions!

Vladimir Poroikov
Corresponding Member of the Russian Academy of Sciences, Prof. Dr.

Roman Efremov
Prof. Dr.

Scientific Program

Tuesday May 24, 2022			
Chairpersons: Vladimir Poroikov, Roman Efremov			
10:00	10:20	Opening of the Symposium	
Plenary lecture			
10:20	11:00	Frank Eisenhaber	THE TARGET IDENTIFICATION BOTTLENECK
Oral presentations			
11:00	11:20	Shuguang Yuan	FROM GPCR BASIC RESEARCH TO DRUG DISCOVERY VIA COMPUTATIONAL METHODS
11:20	11:40	Dmitry Shulga	THE PHILOSOPHY AND PROSPECTS OF FRAGMENT CONTRIBUTION ESTIMATIONS IN DRUG DISCOVERY
11:40	12:00	S. Balaji	INTERACTION AND INHIBITION OF ALPHA-GLUCOSIDASE WITH SELECTED MONOTERPENES
Keynote lectures			
12:00	12:30	Elena Ponomarenko	TARGETED PROTEOMICS FOR HEALTH ANALYTICS: OPPORTUNITIES AND CHALLENGES
12:30	13:00	Alexander Kel	GENE NETWORKS AND DRUGS. WHAT CAN WE LEARN USING BIO- AND CHEMOINFORMATICS?
Oral presentations			
13:00	13:20	Vladimir Ivanisenko	ANDSYSTEM AUTOMATED RECONSTRUCTION OF GENE NETWORKS FOR OMICS-DATA INTERPRETATION IN MEDICAL AND BIOLOGICAL RESEARCH
13:20	13:40	Sajjad Gharaghani	SYSTEM PHARMACOLOGY IN DRUG DISCOVERY
13:40	14:00	Eugene Radchenko	DISCOVERY OF NOVEL TANKYRASE INHIBITOR CHEMOTYPES AN INSIGHTFUL TEST CASE FOR VIRTUAL SCREENING AND MOLECULAR MODELING APPROACHES
Lunch break 14:00-16:00			

Chairpersons: Hanoch Senderowitz, Maria Khrenova			
Plenary lectures			
16:00	16:30	Hanoch Senderowitz	COMPUTATIONAL STUDIES ON GREEN PESTICIDES
16:30	17:00	Dmitry Ivankov	ALPHAFOLD: PREDICTS OR RECOGNIZES THE PROTEIN STRUCTURE?
Oral presentations			
17:00	17:20	Guzel Minibaeva	DE NOVO GENERATION OF SYNTHETICALLY FEASIBLE MOLECULES
17:20	17:40	Anna Tashchilova	SYNTHESIS, DOCKING AND IN VITRO ANTICOAGULANT ACTIVITY ASSAY OF RHODANINE DERIVATIVES OF PYRROLO3,2,1-IJQUINOLIN-2(1H)-ONE AS NEW INHIBITORS OF FACTOR XA AND FACTOR XIA
17:40	18:00	Andrey Buglak	QSPR ANALYSIS IN PHOTONICS
Keynote lectures			
18:00	18:30	Walter F. de Azevedo, Jr.	HARNESSING MACHINE LEARNING FOR DRUG DISCOVERY
18:30	19:00	Artem Cherkasov	THE USE OF DEEP DOCKING FOR AUTOMATED, CONSENSUS-BASED HIT IDENTIFICATION IN DRUG DISCOVERY
Oral presentations			
19:00	19:20	Leonid Stolbov	SELF CONSISTENT CLASSIFIER SAR APPROACH
19:20	19:40	Miguel Guerrero-Gonzalez	DEVELOPMENT OF THE "VSAfiR" METHOD AND ITS APPLICATION IN THE DEVELOPMENT OF ANTIEPILEPTICS

Wednesday May 25, 2022			
Chairpersons: Kunal Roy, Timur Madzhidov			
Plenary lectures			
10:00	10:30	Weiliang Zhu	MOLECULAR DYNAMICS STUDIES ON THE INTERACTIONS BETWEEN SARS-COV-2 SPIKE PROTEIN AND HACE2 OR MABS
10:30	11:00	Dmitry Osolodkin	COMPETITION AND COLLABORATION OF IN SILICO AND IN VITRO SCREENING IN THE SEARCH FOR NEW ANTIVIRAL COMPOUNDS
Oral presentations			
11:00	11:20	Igor Polyakov	SARS-COV-2 MAIN PROTEASE INHIBITION WITH CARMOFUR A COMPUTATIONAL STUDY
11:20	11:40	Tugba Taskin-Tok	MOLECULAR DOCKING-ASSISTED INVESTIGATION OF CU(II) COMPLEXES CARRYING “SNS” PINCER-TYPE PYRIDINE-THIOETHER LIGANDS AS POTENTIAL DRUG CANDIDATES AGAINST SARS-COV-2
11:40	12:00	Elena Aliper	A STRUCTURAL-DYNAMIC MODEL OF SARS-COV-2 SPIKE TRANSMEMBRANE DOMAIN IN CONJUNCTION WITH THE HR2 REGION. IMPLICATIONS FOR MEMBRANE FUSION
Keynote lectures			
12:00	12:30	Kyoung Tai No	DRUG DISCOVERY WITH FRAGMENT MOLECULAR ORBITAL (FMO)
12:30	13:00	Maria Khrenova	HOW DO ENZYMES RECOGNIZE SUBSTRATES AND INHIBITORS: STRUCTURAL AND ELECTRON DENSITY ASPECTS
Oral presentations			
13:00	13:20	Artem Kniga	COMPUTATIONAL CHARACTERIZATION OF N-ACETYLASPARTYLGLUTAMATE SYNTHETASE FROM THE PROTEIN PRIMARY SEQUENCE TO PLAUSIBLE CATALYTIC MECHANISM
13:20	13:40	Anastasia Fomina	ACTIVITY PREDICTION OF SARS-COV-2 MPRO INHIBITORS BASED ON ENSEMBLE DOCKING AND MACHINE LEARNING
13:40	14:00	Anton Chugunov	PHF10 THE SUBUNIT OF PBAF CHROMATIN REMODELING COMPLEX STRUCTURE AND FUNCTION PREDICTIONS
Lunch break 14:00-16:00			
Chairpersons: Pavel Polishchuk, Vladimir Palyulin			

Young scientists flash presentations			
16:00	16:10	Dessiree Allyssa Tina	UNVEILING THE POTENTIAL DRUG LIGANDS AGAINST VIRULENCE-RELATED HYPOTHETICAL PROTEIN IN CRYPTOCOCCUS NEOFORMANS AN *IN SILICO* ANALYSIS APPROACH
16:10	16:20	Hadiatullah Hadiatullah	VIRTUAL SCREENING OF PLANT-DERIVED COMPOUNDS TARGETING HYDROLYTIC AND LIGNIN DEGRADING ENZYMES OF GANODERMA BONINENSE
16:20	16:30	Anatoliy Bulygin	COMPUTATIONAL APPROACH FOR IMPROVING OF KNOWN PERSPECTIVE SARS-COV-2 MPRO INHIBITORS
16:30	16:40	Debanjan Saha	MULTI-TARGET APPROACH ON LEISHMANIA DONOVANI AND FINDING OUT POTENT INHIBITORS FOR ESSENTIAL ENZYMES
16:40	16:50	Alexandra Sadovskaya	SMMOLE - PIPELINE FOR SEARCHING BIOLOGICAL PROPERTIES OF SECONDARY METABOLITES BASED ON THEIR MOLECULAR STRUCTURES
16:50	17:00	Mohammed Efendi	TESTING THE ACTIVITY OF BIGUANIDES AND SOME NOVEL DESIGNED MOLECULES AGAINST SARS-COV-2 PROTEINS, IN SILICO STUDY
17:00	17:10	Egor Kozlov	DIMERIC STATES OF TRANSMEMBRANE SEGMENTS OF THE DDR1 RECEPTOR PREDICTED BY ATOMISTIC MODELING
17:10	17:20	Arkaprava Banerjee	APPLICATION OF 2D-QSAR AND CHEMICAL READ-ACROSS ALGORITHM TO PREDICT THE ANDROGEN RECEPTOR BINDING AFFINITY
17:20	17:30	Nadezhda Biziukova	NFORMATION EXTRACTION FROM TEXTS ANTIVIRAL AGENTS ACTIVE AGAINST VIRUS OR HOST PROTEINS
17:30	17:40	Aleksandra Ivanova	STRUCTURAL OPTIMIZATION OF TUBULIN INHIBITORS
17:40	17:50	Maksim Perfilev	THE CONSENSUS ENSEMBLE NEURAL NETWORK MULTITARGET MODEL OF ANXIOLYTIC ACTIVITY
17:50	18:00	Alina Kutlushina	MOLECULAR DYNAMIC PHARMACOPHORE AND ITS APPLICATION IN DESIGNING NOVEL MARK4 INHIBITORS
18:00	18:10	Ivan Kuznetsov	ALINA - A DEEP LEARNING BASED PROGRAM FOR PREDICTION OF RNA SECONDARY STRUCTURE WITHOUT SPECIFICATION OF THERMODYNAMIC PARAMETERS
18:10	18:20	Ana Luisa Chavez-Hernandez	TOWARDS THE DE NOVO DESIGN OF HIV-1 PROTEASE INHIBITORS BASED ON NATURAL PRODUCTS
18:20	18:30	Gabriela Bitencourt-Ferreira	EXPLORING THE SCORING FUNCTION SPACE FOR STRUCTURE-BASED DRUG DESIGN
18:30	18:40	Luis Heriberto Vazquez Mendoza	REPURPOSING OF FDA-DRUGS AS POTENTIAL ERB AGONISTS USING MULTICOMPLEX-BASED

			PHARMACOPHORE MAPS. A NEW APPROACH IN BREAST CANCER THERAPY
18:40	18:50	Alessandra Latorre	IN SILICO DESIGN OF QUERCETIN DERIVATIVES WITH POTENTIAL DUAL INHIBITORY ACTIVITY AGAINST GSK3 AND CDK5P25 FOR THE TREATMENT OF ALZHEIMER'S DISEASE
18:50	19:00	Edgar López-López	CONSENSUS VIRTUAL SCREENING OF NATURAL PRODUCT DERIVATIVES AGAINST TUBULIN

Thursday May 26, 2022			
Chairpersons: Athina Geronikaki, Dmitry Osolodkin			
Plenary lectures			
10:00	10:30	Garikapati Narahari Sastry	THE STATUS OF THE COMPUTER-AIDED DRUG DESIGN: THEN, NOW AND FUTURE
10:30	11:00	Roman Efremov	COMPUTATIONAL DRUG DESIGN FOR MEMBRANE TARGETS: DIVING INTO COMPLEX DETAILS
Oral presentations			
11:00	11:20	Petr Popov	SPATIOTEMPORAL IDENTIFICATION OF BINDING SITES WITH COMPUTER VISION
11:20	11:40	Vladimir Sulimov	NEW INHIBITORS OF THE COAGULATION FACTOR XIIA DOCKING AND EXPERIMENTAL VERIFICATION
11:40	12:00	Anastasia Borovik	AN INSIGHT INTO THE ORIGIN OF MICROTUBULE-CURLING EFFECT OF PODOPHYLLOTOXIN ESTERS MOLECULAR DYNAMICS STUDY
Keynote lectures			
12:00	12:30	Timur Madzhidov	CONDENSED GRAPH OF REACTION - SWISS-KNIFE TOOL FOR REACTION INFORMATICS
12:30	13:00	Kunal Roy	CHEMICAL READ-ACROSS PREDICTIONS OF ECOTOXICITY DATA
Oral presentations			
13:00	13:20	Yuriy Orlov	DEVELOPMENT OF BIOMEDICAL EDUCATIONAL PROGRAMS
13:20	13:40	Kuppusamy Selvam Mukunthan	A COMPREHENSIVE COMPUTATIONAL PHARMACOKINETICS IDENTIFICATION OF BIOTRANSFORMED LEADS FROM CURCUMA CAESIA ROXB
13:50	14:00	Pavel Pogodin	TCSTF, TOOL FOR CATEGORIZATION OF SHORT TEXT FRAGMENTS
Lunch break 14:00-16:00			
Chairpersons: Artem Cherkasov, Alexey Lagunin			
Keynote lectures			
16:00	16:30	Olga Bocharova	PHARMACOLOGICAL POTENTIAL OF MULTIPHYTOADAPTOGEN AS POLYVALENT MEDICATION: IN SILICO, IN VITRO, IN VIVO AND CLINICAL STUDIES
16:30	17:00	Marcus Scotti	NATURAL PRODUCTS DATABASES AS VALUABLE SOURCES OF BIOACTIVE STRUCTURES FOR VIRTUAL SCREENING

Oral presentations			
17:00	17:20	Pavel Vassiliev	THE CONSENSUS ENSEMBLE MULTIDESCRIPTOR MULTITARGET NEURAL NETWORK MODELING OF PHARMACOLOGICAL ACTIVITY OF CHEMICAL COMPOUNDS
17:20	17:40	Evgenia Alimbarashvili	DATABASE OF ANTIMICROBIAL ACTIVITY AND STRUCTURE OF PEPTIDES (DBAASP) - FINDING A WAY OUT OF MICROBIAL RESISTANCE
Keynote lectures			
17:40	18:10	Oxana Galzitskaya	AMYLOIDOGENIC PEPTIDES NEW CLASS OF ANTIMICROBIAL PEPTIDES WITH THE NOVEL MECHANISM OF ACTIVITY
18:10	18:40	Dmitry Filimonov	SIMILARITY ASSESSMENTS IN DRUG DISCOVERY
Plenary lecture			
18:40	19:20	José Medina-Franco	CHEMOINFORMATICS IN DRUG DISCOVERY AND PUBLIC HEALTH: PROGRESS AND CHALLENGES AHEAD
19:20	20:00	Closure of the XXVIII Symposium on Bioinformatics and Computer-Aided Drug Discovery	

PLENARY/KEYNOTE LECTURES

PHARMACOLOGICAL POTENTIAL OF MULTIPHYTOADAPTOGEN AS POLYVALENT MEDICATION: *IN SILICO, IN VITRO, IN VIVO* AND CLINICAL STUDIES

***O. Bocharova*[1], *N. Ionov*[4], *I. Kazeev*[1], *V. Shevchenko*[1], *E. Bocharov*[1], *R. Karpova*[1], *O. Sheychenko*[2], *V. Kucheryanu*[3], *V. Kosorukov*[1], *V. Matveev*[1], *D. Filimonov*[4], *A. Lagunin*[4], *V. Poroikov*[4]**

[1]*N.N. Blokhin National Medical Research Center of Oncology, Moscow, Russia*
[2]*All-Russian Scientific Research Institute of Medicinal and Aromatic Plants, Moscow, Russia*
[3]*Institute of General Pathology and Pathophysiology, Moscow, Russia*
[4]*Institute of Biomedical Chemistry, Moscow, Russia*

Many human diseases including cancer, diabetes, autoimmune disorders, degenerative diseases and others are multi-factorial. In this case, pharmaceutical agents acting on a single target do not provide their efficient curation. Multita.rgeted drugs exhibiting the pleiotropic pharmacological effects have certain advantages due to the normalization of the different etiology complex pathological processes. Medicinal plants extracts containing multiple phytoconstituents are widely used in the traditional medicines for treatment of the multi-factorial disorders [1]. Experimental studies of multicomponent compounds' mixtures pharmacological potential are very expensive and time-consuming. *In silico* evaluation of multi-component medications pharmacological potential may not only provide the basis for selecting the most promising directions of testing but also for identifying potential additive/synergistic effects. Such estimations could be obtained using computer programs PASS and PharmaExpert [2].

Multiphytoadaptogen (MPhA) containing seventy major phytocomponents of different chemical classes from forty extracts of medicinal plants is investigated in preclinical and clinical studies for about thirty years [3, 4]. Antitumor action against adenocarcinoma of the ovaries and cervix as well as against hypernephroma has been found in *in vitro* experiments. Experiments *in vivo* on high-cancer CBA mice-males have demonstrated potent MPhA effect against hepatocellular carcinoma. *In vivo* a 100% antimetastatic effect was found on lung carcinoma. Antitumor activity in stage four advanced gastric cancer has been shown in the clinics. Chemopreventive (oncoprophylaxis) effect was demonstrated in CBA mice, as well as in the clinics for the treatment of precancerous oral leukoplakia. The therapeutic effects of age-related diseases - benign prostatic hyperplasia and Parkinson's disease were revealed. In addition, neuroprotective as well as antioxidant, antimutagenic, radioprotective, immunomodulating MPhA action have been found.

Significant chemical diversity of the MPhA components may provide a great biological activity profile of the multicomponent preparation, which could be estimated using PASS and PharmaExpert. We have analyzed the PASS estimations for the separate phytocomponents and show that most of the predicted antitumor and antimetastatic effects correspond to those found *in vitro*, *in vivo* and in clinics. For more than 50 compounds antineoplastic effect is predicted against the thirteen types of tumors. Antimutagenic, immunomodulating, radioprotective, neuroprotective and antiparkinsonian effects have been predicted for some phytocomponents as well. Four probable mechanisms of action are revealed, including transcription factor NF kappa B inhibitor; caspase 3 stimulant; apoptosis agonist; antioxidant, which also corresponds to the experimental results. Integration of the predictions for separate phytocomponents using PharmaExpert indicates the probable additive/synergistic actions against tumors of different etiology.

1. Li, F.S., Weng, J.K. (2017) *Nature Plants*, 3, 17109.
2. Lagunin, A.A., Goel, R.K., Gawande, D.Y. et al. (2014) *Natural Product Reports*, 31, 1585.
3. Bocharova, O., Karpova, R., Bocharov, E. et al. (2020) *Russian Journal of Biotherapy*, 19, 22.
4. Bocharova, O., Matveev, V., Bocharov, E. et al. (2021) *Russian Journal of Biotherapy*, 20, 42.

THE USE OF DEEP DOCKING FOR AUTOMATED, CONSENSUS-BASED HIT IDENTIFICATION IN DRUG DISCOVERY

A. Cherkasov

The University of British Columbia, Vancouver, Canada

Recent explosive growth of "make-on-demand" chemical libraries brings unprecedented opportunities but also raises significant challenges to the field of computer-aided drug discovery. To address this expansion of the accessible chemical universe, molecular docking needs to accurately rank tens of billions of chemical structures, calling for the development of automated hit-selecting protocols to minimize human intervention and error.

We have recently developed an artificial intelligence-driven virtual screening pipeline that utilizes Deep Docking with Autodock GPU, Glide SP, FRED, ICM and QuickVina2 programs to screen 40 billion molecules against SARS-CoV-2 main protease (Mpro) [1]. This campaign returned a significant number of experimentally confirmed inhibitors of Mpro enzyme, and also enabled to benchmark the performance of twenty-eight hit-selecting strategies of various degrees of stringency and automation. These findings provide new starting scaffolds for hit-to-lead optimization campaigns against Mpro and encourage the development of fully automated end-to-end drug discovery protocols integrating machine learning and human expertise.

This work was funded by Canadian Institutes of Health Research (CIHR), Canadian 2019 Novel Coronavirus (2019-nCoV) Rapid Research grants (OV3-170631 and VR3-172639), and generous donations for COVID-19 research from TELUS, Teck Resources, 625 Powell Street Foundation, Tai Hung Fai Charitable Foundation, Vancouver General Hospital Foundation. The author also thanks the Dell Technologies HPC and AI Innovation Lab for their support and partnership in providing the HPC platform (PowerEdge servers) to accelerate the AI algorithms, and the UBC Advanced Research Computing team for providing access and technical support for the Sockeye supercomputing cluster.

1. Gentile, A., Fernandez, M., Ban, F. et al. (2021) *Chem. Sci.*, 2021, 12, 15960.

COMPUTATIONAL DRUG DESIGN FOR MEMBRANE TARGETS: DIVING INTO COMPLEX DETAILS

***R. Efremov*[1-3]**

[1]*Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Moscow, Russia*
[2]*National Research University Higher School of Economics, Moscow, Russia*
[3]*Moscow Institute of Physics and Technology (State University), Dolgoprudny, Russia*

Membrane proteins (MPs) make up about 30% of the proteome of humans and most other organisms, but only ~2% of them have a known spatial structure. MPs are involved in the most important processes in the cell, including signal and energy transduction, molecular and ion transport, enzymatic reactions, etc. MPs are the most important pharmacological targets, since more than half of all known drugs act on these proteins. For the rational drug design and therapies for the most severe diseases, such as cancer, cardiovascular, neurodegenerative and infectious, it is necessary to have atomistic models of the structural-dynamic organization of MPs. Along with rapidly developing experimental methods (cryo-EM, NMR, X-ray, SAXS, etc.), methods of computational experiment provide invaluable assistance.

When constructing MPs' models and studying the molecular mechanisms of their functioning, both standard computer modeling methods that have proven themselves well in the case of water-soluble proteins and special approaches that take into account important features of MPs are used. The first include methods of structural bioinformatics, homology modeling, sampling of the proteins phase space via molecular dynamics (MD) / Monte Carlo, molecular docking, etc. These approaches provide adequate information for drug design, especially if the active center is located outside the membrane. The task is much more difficult if the biologically active compound binds and acts in the transmembrane domain (TMD) of the protein. So, in recent years it has become obvious that, in this case, it is necessary to take into account a number of features that were previously practically not considered. Among them: (1) The degree of autonomy of the protein TMD, i.e. whether it is necessary to take into account the juxtamembrane and more distant domains that can be allosterically coupled; (2) Heterogeneity (mosaic) and complex dynamics of the surrounding layer of lipids, water and ions, forming for each MP type its unique "dynamic molecular portrait"; (3) The role of individual lipids in maintaining the spatial structure of MP and transitions between its functional states; (4) Contribution of TMD homo- and hetero-oligomerization processes.

This work presents a comprehensive computational approach to modeling MPs of several classes (receptor tyrosine kinases, ion channels, viral proteins) – potentially important pharmacological targets. The consisted application of both standard modeling methods (MD, homology modeling, ensemble docking, etc.) and original approaches (detailed mapping of spatio-temporal characteristics and physico-chemical properties of MPs and membranes, analysis of key physical interactions in these systems, etc.) is described. It is concluded that the successful computer-aided design of drugs acting on target proteins in cell membranes requires consideration of the MP/lipids/water/ions system as a single whole, and the relationships between all components should be taken into account as carefully as possible.

This work was supported by RSF grant No. 18-14-00375.

THE TARGET IDENTIFICATION BOTTLENECK: DECLINE IN BIOMOLECULAR MECHANISM DISCOVERY AFTER 2000

F. Eisenhaber

Genome Institute of Singapore and Bioinformatics Institute, Singapore

Rational drug design (or generally that of medical interventions) requires insight into mechanism in physiology and disease, desirably at the level of bimolecular mechanisms. This is the common theme in all success cases.

It is generally believed that full human genome sequencing was a watershed event in human history that boosted biomedical research, biomolecular mechanism discovery and life science applications. At the same time, researchers in the field of genome annotation see that there is a persisting, substantial body of functionally insufficiently or completely not characterized genes (for example, ~10,000 protein-coding in the human genome) despite the availability of full genome sequences. A survey of the biomedical literature shows that the number of reported new protein functions had been steadily growing until 2000 but the trend reversed to a dramatic decline thereafter (1,2) when, at the same time, the annual amount of new life science publications doubled between 2000 and 2017.

This reduction in the supply of newly characterized pathways has profound implications for the drug development pipelines in industry as well as for research. The example of the SUGCT function discovery (3) shows the significance that many uncharacterized genes will have in aging, metabolic diseases and their complications.

1. Sinha, S., Eisenhaber, B., Jensen, L.J., Kalbuaji, B., Eisenhaber, F. (2018) *Proteomics*, 18(21-22), e1800093.
2. Eisenhaber, F.J. (2012) *Bioinform. Comput. Biol.*, 10(5), 1271001.
3. Niska-Blakie, J., Gopinathan, L., Low, K.N., Kien, Y.L., Goh, C.M.F., Caldez, M.J., Pfeiffenberger, E., Jones, O.S., Ong, C.B., Kurochkin, I.V., Coppola, V., Tessarollo, L., Choi, H., Kanagasundaram, Y., Eisenhaber, F., Maurer-Stroh, S., Kaldis, P. (2019) *Cell Mol. Life Sci*., 77, 3423–3439.

SIMILARITY ASSESSMENTS IN DRUG DISCOVERY

D. Filimonov*, *V. Poroikov

Institute of Biomedical Chemistry, Moscow, Russia

Assessment of molecular similarity is one of the popular computational methods used in medicinal chemistry. It is widely used for design and synthesis of analogs during the drug optimizations process, and for an umbrella patenting of pharmaceutical agents from certain chemical series with a particular biological activity. Computational tools based on the similarity estimation are implemented in large chemical libraries providing the search for analogs with the expected similar properties.

The hypothesis that structurally similar compounds exhibit similar biological effects or some other properties is taken as an axiom. However, molecular similarity, as a paradigm, contains many implicit and explicit assumptions. One does not know *a priory* which properties of the molecular structure are essential for its biological activity; therefore, the description of the structure can be only heuristic [1]. The selection of molecular descriptors and the estimation of molecular similarity based on this selection crucially determine the final result of the study [2, 3]. However, for novel pharmacological targets (like SARS-CoV-2 coronavirus proteins), when only limited number of antiviral agents that may be used as a "query" are known, similarity assessment is the method-of-the-choice.

The biological activity of drug-like compound mostly expresses the result of molecular recognition, which, in turn, depends primarily on the correspondence between specific ligand atoms and the pharmacological target. On this basis, we have proposed an enhancement of the similarity principle, called the “local correspondence concept”. Based on this concept, we have developed a consistent system of descriptors named Multilevel and Quantitative Neighborhoods of Atoms (MNA and QNA, respectively). MNA descriptors are successfully used for predicting biological activity spectra of drug-like molecules in the PASS software [2, 4] for about 30 years. For QSAR/QSPR modeling we have proposed a novel QNA based "Star Track" approach [5], where, in accordance with the local correspondence concept, any molecule is represented as a set of points in the two-dimensional space of QNA descriptors.

The MNA and QNA descriptors have the form of disordered sets, which significantly distinguishes them from the other molecular descriptors. MNA descriptors are the strings of symbols (linear notations of atoms with their surroundings). QNA descriptors are presented by the pairs of real numbers, P and Q, for each atom of the molecule. For MNA descriptors, a well-known Tanimoto measure of similarity of two discrete sets may be used. The peculiarity of the QNA descriptors does not allow the use of both set-theoretic similarity measures and known similarity measures based on distances due to the disordered nature of QNA descriptors. We have proposed a similarity assessment using the QNA descriptors, in which the Todeschini approach [6] and the Tanimoto similarity are used. This approach is implemented in the web service freely available via the Internet, which provides the computer-aided search for novel anticoronavirus agents [7].

Challenges and opportunities of molecular similarity application in drug discovery including the case studies based on our experience will be considered.

The study was supported in the framework of the Russian Foundation for Basic Research grant No. 20-04-60285.

1. Bender, A. (2010) *Expert Opin. Drug Discov.*, 5, 1141.
2. Muratov, E.N., Bajorath, J., Sheridan, R.P. et al. (2020) *Chemical Society Reviews*, 49, 3525.
3. Bragina, M.E., Daina, A., Perez, M.A.S. et al. (2022) *Int. J. Mol. Sci.*, 23, 811.
4. Poroikov, V.V., Filimonov, D.A., Gloriozova, T.A. (2019) *Russian Chemical Bulletin*, 68, 2143.
5. Filimonov, D.A., Zakharov, A.V., Lagunin, A.A. et al. (2009) *SAR & QSAR Environ. Res.*, 20, 679.
6. Mauri, A., Ballabio, D., Todeschini, R. et al. (2016) *J. Cheminform.*, 8, 49.
7. Anticoronavirus activity prediction URL [http://www.way2drug.com/anticovid/#services], accessed on April 22th, 2022.

ALPHAFOLD: DOES IT PREDICT OR RECOGNIZE THE PROTEIN STRUCTURE?

***D.N. Ivankov*[1], *M.A. Pak*[1], *A.V. Finkelstein*[2,3,4]**

[1]*Skolkovo Institute of Science and Technology, Moscow, Russia*
[2]*Institute of Protein Research, Russian Academy of Sciences, Pushchino, Moscow Region, Russia*
[3]*Biotechnlogy Department, Lomonosov Moscow State University, Pushchino, Moscow Region, Russia*
[4]*Biology Department, Lomonosov Moscow State University, Moscow, Russia*

Prediction of three-dimensional (3D) protein structure from the protein amino acid sequence has been a long-standing problem in protein physics and structural bioinformatics for more than 50 years [1]. Two years ago, a deep neural network program AlphaFold won Critical Assessment of Protein Structure Prediction (CASP), a biannual competition on blind prediction of protein 3D structure from the sequence [2]. More importantly, AlphaFold models were of near-to-experimental accuracy [3]. This enormous success poses a question: does AlphaFold really "predict" protein 3D structure or just recognize it by similarity with already known 3D structures? In other words, has AlphaFold learned the free energy potential functions, or it bases mostly on analogy with 3D protein structures and sequences that are already stored in the databanks?

To check a putative potential function of AlphaFold, we first asked if AlphaFold could estimate the effect of mutations on protein stability, which is equivalent to the task of ranking two protein structures based on their stabilities. The only way to do this test was to use AlphaFold's level of confidence to score different mutant structures: the more confident the prediction, the more stable the corresponding 3D structure. However, we found virtually no correlation between AlphaFold confidence and experimental effect of mutations, which means that the AlphaFold's level of confidence does not use the free energy function in the output [4].

On the other hand, we checked if the statistical power of the databases on which AlphaFold was trained, is enough to make successful predictions. For this, we estimated the expected similarity of a random sequence S to the most similar to it chain S' from the set $\boldsymbol{\Sigma_N}$ of $\boldsymbol{N}$ other random sequences. Actually, we pose a question: Is the set $\boldsymbol{\Sigma_N}$ large enough to include a sequence S', which is so similar to S that their 3D structures are very similar? We found that the Protein Data Bank and UniProt databases contain the sufficient number of sequences to find among them a sequence with ~25% sequence identity to an arbitrary random sequence. On average, such sequence identity corresponds to the superfamily level with structural divergence of only 1.7Å±0.5Å [5]. This level is enough to recognize the topology of the 3D structure and then to refine the recognized 3D structure using conventional methods.

To summarize, we show that the basis for the tremendous success of AlphaFold is a very clever usage of huge protein databases that already cover all or almost all of the protein superfamilies that exist in nature. The AlphaFold team managed to use the deep learning technology for finding the effective way of the recognition of the correct fold, and then to refine it.

AVF acknowledges support from the Russian Science Foundation (grant No. 21-14-00268).

1. Dill, K., MacCallum, J. (2012) *Science*, 338, 1042-1046.
2. Jumper, J. et al. (2021) *Nature*, 596, 583–589.
3. Callaway, E. (2022) *Nature*, 604, 234–238.
4. Pak, M.A. et al. (2021) *bioRxiv*, 460937.
5. Lesk, A.M., Chothia, C. (1986) *Phil. Trans. R. Soc. Lond. A* 317, 345–356.

GENE NETWORKS AND DRUGS. WHAT CAN WE LEARN USING BIO- AND CHEMOINFORMATICS?

D. Stelmashenko[1,2], O. Kel-Margoulis[1], S. Apalko[3], A. Kel[1,2,4]

[1]geneXplain GmbH, Wolfenbüttel, Germany
[2]biosoft.ru, Novosibirsk, Russia
[3]City Hospital №40, St. Petersburg, Russia
[4]Institute of Chemical Biology and Fundamental Medicine, Novosibirsk, Russia

Correct identification of the therapeutic targets is the first step on the way to successful treatment of complex diseases such as cancer, autoimmune diseases, neurogenerative disorders, sepsis. "Gene networks", better to say, networks of signal transduction, gene regulation and protein-protein interactions that are acting in the cells and organs in norm and pathology have clues to identification of effective drug targets. Such molecular networks practically always are characterized by a hierarchical structure when a few important molecules, so called master-regulators, that are positioned at the top of the hierarchy and possess control over activity of hundreds and thousands of genes in all molecular and physiological processes in the cells and in the whole organism. In cases of complex diseases, often, a few pathologically altered master-regulators hijack the gene regulatory networks and lead to the disease state of the system. In cancer, such alterations, so called "driver mutations" often cause a complete rewiring of the gene regulatory network that are characterized by numerous positive feedback loops [1].

Development of effective computational approaches for identification of master-regulators as prospective drug targets based on multi-omics data is one of the challenges addressed by bioinformatics and chemoinformatic today. Our current work aims to develop an algorithm for reconstruction of the molecular mechanism of a certain pathology and selection of effective therapies based on the personalized model of the disease. The algorithm introduced in this work can be applied to analyse any combinations of 5 major types of omics data: genomics, transcriptomis, epigenomics, proteomics and metabolomics. It is called an algorithm of "*Walking pathways*" [1]. The algorithm consists of three main steps: 1) TRANSFAC analysis to identify complexes of transcription factors (TF) that dysregulate gene expression in the disease; 2) TRANSPATH network analysis to identify common regulators of the TFs, identification of master-regulators with feedback loops as potential drug targets; 3) Search in HumanPSD database for known drugs for the found targets as well as usage of chemoinformatics tool PASS to identify new drugs targeting the predicted targets. This algorithm is available at the Genome Enhancer web site (https://genexplain.com/genome-enhancer/).

In this work, we applied Genome Enhancer to analyze data from transcriptomic (RNA-seq) and whole exome sequencing to reconstruct the molecular mechanism of pathology in individual patients with colorectal adenocarcinoma, and identification of their drug targets. Based on the found drug targets we prioritized drugs to identify the potential of the tumor to be sensitive or resistant to the proposed drug treatment. We demonstrated a clear correlation between the sensitivity to the chosen therapy predicted by the algorithm and observed therapy response in the individual cases. So, the target-oriented approach is one of the most effective approaches in personalized medicine.

1. Kel A. et.al. (2019) *BMC Bioinformatics,* 20 (Suppl 4), 119.

HOW DO ENZYMES RECOGNIZE SUBSTRATES AND INHIBITORS: STRUCTURAL AND ELECTRON DENSITY ASPECTS

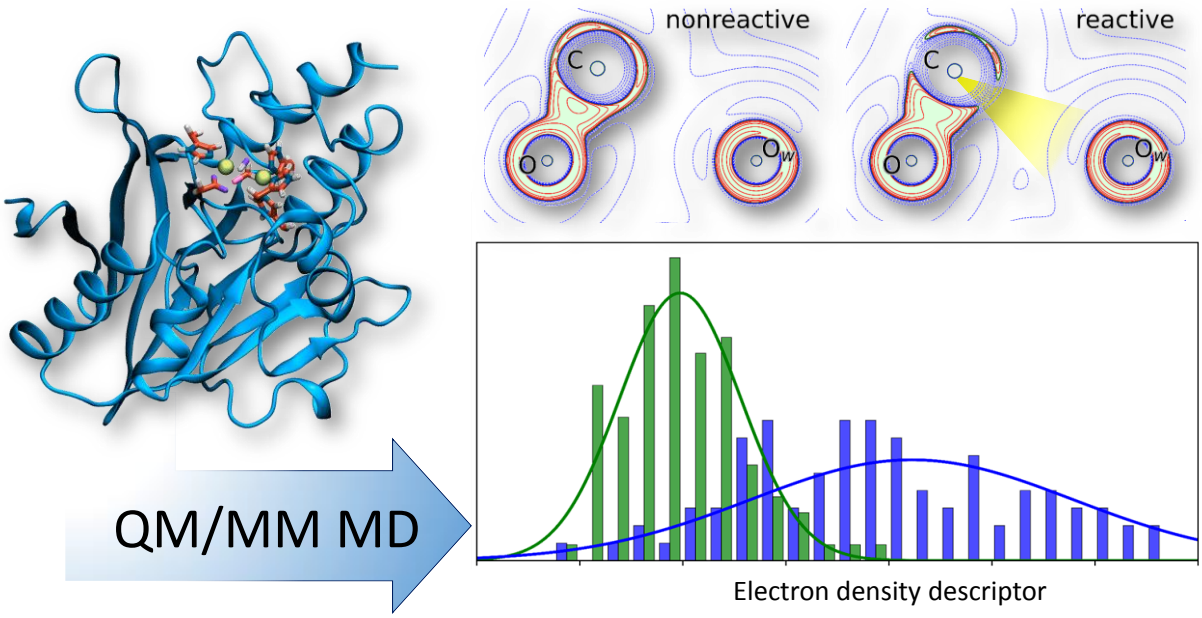
M. Khrenova[1,2]

[1]*Lomonosov Moscow State University, Moscow, Russia*
[2]*Federal Research Centre "Fundamentals of Biotechnology", RAS, Moscow, Russia*

Tremendous experimental efforts are made to accumulate data on substrate specificity of various enzymes. There are also numerous studies in the literature on the systematization of the obtained data based on bioinformatic analysis, as well as on the prediction of substrates preferred for the selected enzymes based on big data analysis. Mainly, such analysis is based on the search of the complementary regions of substrates and binding sites of enzymes. Importantly, amino acid residues forming the active sites are the same within the selected class. Thus, bioinformatic analysis can answer questions related to the complementarity of peptide substrates and protease binding sites, which certainly affects their reactivity (the value of the catalytic constant k_{cat} of the stationary Michaelis-Menten kinetics). However, these methods cannot explain the reasons underlying different reactivity (manifested in a change in k_{cat}) with respect to different substrates, which is of interest to fundamental science.

It is possible to determine the origin of substrate specificity within the framework of molecular modeling based on quantum theory, which makes it possible to quantitatively describe and characterize interatomic and intermolecular interactions in the active sites of enzymes. Herein, we present results of the study of the dynamic behavior of enzyme-substrate complexes using the molecular dynamics method with combined quantum mechanics/molecular mechanics potentials, followed by quantitative analysis of the interactions of the enzyme and the substrate and characterization of substrate activation by the enzyme using electron density descriptors.

This study was supported by the grant of the President of Russian Federation for young doctor of sciences (MD-1390.2022.1.4).



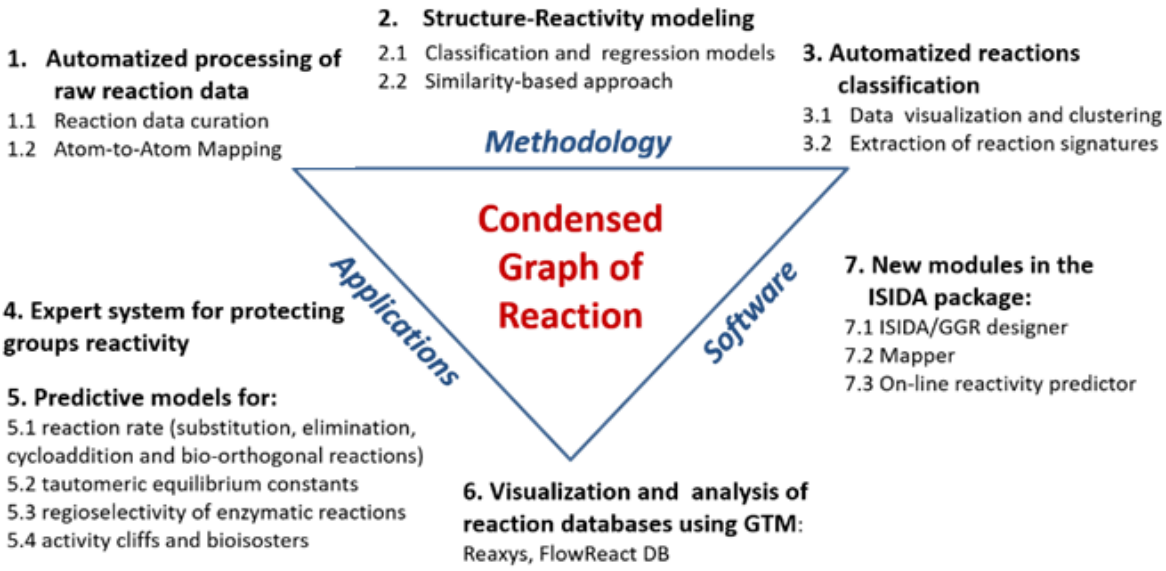
CONDENSED GRAPH OF REACTION - SWISS-KNIFE TOOL FOR REACTION INFORMATICS

***T. Madzhidov*[1], *A. Rakhimbekova*[1], *V. Afonina*[1], *A. Fatykhova*[1], *D. Zankov*[1], *A. Varnek*[2]**

[1]*Kazan Federal University, Kazan, Russia*
[2]*University of Strasbourg, Strasbourg, France*

Wealth of data on chemical compounds and organic reaction has been collected due to increasing performance of synthetic work. The greatest databases contain about 100 mln data on chemical reactions. It opened the door for applications of big data analytics and data mining technologies (including artificial intelligence approaches) in synthetic chemistry [1]. On the other side, it requires that a synthetic chemist has to be very productive and efficient, especially in industrial settings. He has to develop a robust and efficient pathway to the desired compound, select optimal conditions, foresee and avoid possible selectivity issues. At the same time, none of the chemists can follow all of the knowledge generated in the field. Thus, big data analytics and machine learning techniques (including artificial intelligence approaches) attract much attention [1].

Here, we review our efforts on development of approaches reaction data analysis, model building and synthesis prediction. The proposed approach is based on Condensed Graph of Reaction (CGR) technique that allows to encode the reaction equation into one sole molecular graph. The advantage of the CGR technique is that it explicitly encodes reaction center. Also, it is a single pseudo-molecular graph with specific bond types corresponding to formed/broken/changed bonds, and thus it can be manipulated as regular molecules by well-developed chemoinformatics algorithms. The most important is that chemical descriptors can be calculated and applied for reaction characteristics modeling, optimal condition assessment and reaction search enhancements. Also, some task-specific modeling techniques were proposed to take into account reaction complexity and specifics. Thus, we proposed conjugated QSA/PR modeling technique to take into account known chemical laws and relations. We proposed an approach for taking into account conformational lability of molecules based on multi-instance approach. The latter were used for prediction of enantioselectivity of catalysts.



The reviewed developments was supported in the framework of Russian Science Foundation grants (14-43-00024, 14-43-00024П, 19-73-10137), state assignment for science for Kazan Federal University (agreement No 075-03-2021-299/6), Ministry of Science and Higher Education of the Russian Federation (agreement no. 14.587.21.0049, unique identifier RFMEFI58718X0049).

1. Baskin, I.I., Madzhidov, T.I., Antipin, I.S., Varnek, A. (2017) *Russ. Chem. Rev.*, 86, 1127.

CHEMOINFORMATICS IN DRUG DISCOVERY AND PUBLIC HEALTH: PROGRESS AND CHALLENGES AHEAD

J.L. Medina-Franco

DIFACQUIM research group, School of Chemistry, National Autonomous University of Mexico, Mexico City, Mexico

The talk is divided into three main sections. First, we present an introduction to chemoinformatics, mentioning a formal definition, similarities, and differences with other theoretical chemistry disciplines and their broad applicability in research, including drug discovery [1]. After the brief background, we discuss the progress of chemoinformatics applications in a drug discovery program ongoing in our research group focused on epigenetic drug discovery with emphasis on the development of inhibitors of DNA methyltransferases. As part of the results, we will present the current trends of machine learning models generated based on large public compound databases annotated with biological activity and implemented in a free webserver to advance epigenetic drug discovery further [2]. We will also discuss the perspectives of this program that includes the development of poly-epigenetic drug candidates. In the third section of the talk, we will outline significant challenges that, in the author's view, faces chemoinformatics and computer-aided drug design in general. For the discussion's purposes, such challenges are organized into three major categories: those associated with the effective exploration and expansion of the chemical and biological spaces, methodological challenges, and hurdles related to effective communication among research teams, data sharing, education, and training [3].

1. López-López, E. et al. (2021) *J. Chem. Inf. Model*, 61, 26-35.
2. Sánchez-Cruz, N. et al. (2021) *J. Med. Chem.,* 64 12, 8208-8220.
3. Medina-Franco, J.L. (2021) *Front. Drug. Discov.*, 1, 728551.

DRUG DISCOVERY WITH FRAGMENT MOLECULAR ORBITAL(FMO)

H. Lim[1,3], H.N. Jeon[2,3], J.W. Kim[3], K.T. No[1,3,4]

[1]Graduare School, Integrative Biotechnology & Translational Medicine, Yonsei University, Incheon, Republic of Korea
[2]Department of Biotechnology, Yonsei University, Seoul, Republic of Korea
[3]Bioinformatics and Molecular Design Research Center, Incheon, Republic of Korea
[4]Baobab AiBIO Inc., Incheon, Republic of Korea

Empirical force fields, which have been used in drug development for nearly half a century, have made a lot of progress during that time, greatly increasing the reliability of Computer-Aided Drug Design. Nevertheless, the structural and energetic accuracy of the calculated results contain significant errors and are often inconsistent with the experimental results. Ideally, the best method would be a combination of MO and MD, but this would not be feasible with today's classic computers. Fragment MO (FMO) is considered to be the optimal method to calculate the binding of target protein and ligand to the level of post HF MO in the current computational power limit of the computer.

Kitaura et al. developed the fragment molecular orbital (FMO) method for large molecules and molecular systems in 1999 [1], and pair interaction energy decomposition analysis (PIEDA) in 2007 [2].

It has been successfully applied to investigate the protein-ligand interactions based on PIEDA qualitatively and quantitatively [3, 4].

For the past 5 years, our laboratory has obtained good research results by introducing FMO to the characterization of PPIs and the design of inhibitors. In this study, I would like to introduce the FMO-based PPI researches [5-8] that has been conducted so far.

Finally, I will briefly discuss on the possibility of FMO computation using quantum computing.

1. Kitaura, K., Ikeo, E., Asada, T., Nakano, T., Uebayasi, M. (1999) *Chemical Physics Letters*, 313, 701–706.
2. Fedorov, D. G., Kitaura, K. (2007) *Journal of Computational Chemistry* 28, 222–237.
3. Heifetz, A. et al. (2016) *Journal of medicinal chemistry*, 59, 4352–4363.
4. Otsuka, T., Okimoto, N., Taiji, M. (2015) *Journal of Computational Chemistry* 36, 2209–2218.
5. Lim, H., Jin, X., Kim, J., Hwang, S., Shin, K.B., Choi, J., Nam, K.-Y., No, K.T. (2020) *Comput. Struct. Biotechnol. J.*, 17, 1217–1225.
6. Lim, H., Chun, J., Jin, X., Kim, J., Yoon, J.H., No, K.T. (2019). *Sci. Rep.*, 9, 16727.
7. Lim, H., Baek, A., Kim, J., Kim, M.S., Liu, J., Nam, K.-Y., Yoon, J.H., No, K.T. (2020) *Sci. Rep.*, 10, 16862.
8. Lim, H., Jeon, H.N., Rhee, J.K., Oh, B., No, K.T. Quantum computational study of chloride ion attack on chloromethane for chemical accuracy and quantum noise effects with UCCSD and k-UpCCGSD ansatzes, *Scientific Reports*, in press.

COMPETITION AND COLLABORATION OF *IN SILICO* AND *IN VITRO* SCREENING IN THE SEARCH FOR NEW ANTIVIRAL COMPOUNDS

D.I. Osolodkin*, *L.I. Kozlovskaya

[1]*Chumakov FSC R&D IBP RAS (Institute of Poliomyelitis), Moscow, Russia*

Viral infections represent a constant threat for public health and have a potential for unpredictable emergence and expansion to the pandemic scale. The use of small-molecule antiviral drugs is one of the most impotant strategies of prophylaxis and treatment of viral infections, adding to or extending the vaccination programs. Combination of phenotypic and target-oriented screening *in silico* and *in vitro* allows to achieve the maximal efficiency on the early stages of antiviral drug discovery and development. In this presentation, we will present several examples of realisation of this strategy with specific emphasis on performance of both *in silico* and *in vitro* driven approaches.

Analysis of the antiviral activity big data from the publicly available databases with the help of machine learning and dimensionality reduction methods [1-4] allowed us to reveal the new chemotypes of inhibitors of reproduction of tick-borne encephalitis virus (TBEV) and other flaviviruses [4-7], enteroviruses [8], as well as SARS-CoV-2 coronavirus [3, 6, 9]. Antiviral activity of these compounds was confirmed experimentally *in vitro*. We have also developed the *in silico* and *in vitro* methods for studying the interactions of small molecules with viral protein targets, such as TBEV envelope protein E and methyl transferase, and SARS-CoV-2 $3CL^{pro}$ protease [9]. These methods can be applied to study the mechanisms of viral reproduction inhibition, as well as the target-based screening. Structural virology approaches are being employed to study the ligand-protein complexes of the most potent and promising compounds for further optimisation.

Combination of the aforementioned approaches allowed us to establish the integrated workflow of antiviral drug discovery and development, which simplifies the translation of fundamental studies into clinically relevant data. A new compound may enter the workflow on each stage to receive the feedback required for progression in the antiviral pipeline.

This study was supported by the State Research Funding for FSASI "Chumakov FSC R&D IBP RAS" (Institute of Poliomyelitis) (projects No. FNZG-2021-0001 & FNZG-2022-0002).

1. Nikitina, A.A., Orlov, A.A., Kozlovskaya, L.I. et al. (2019) *Database*, 2019, bay139.
2. Sosnina, E.A., Sosnin, S., Nikitina, A.A. et al. (2020) *ACS Omega*, 5, 15039-15051.
3. Horvath, D., Orlov, A., Osolodkin, D.I. et al. (2020) *Molecular Informatics*, 39, 2000080.
4. Orlov, A.A., Khvatov, E.V., Koruchekov, A.A. et al. (2019) *Molecular Informatics*, 38, 1800166.
5. Vasilenko, D.A., Dueva, E.V., Kozlovskaya, L.I. et al. (2019) *Bioorganic Chemistry*, 87, 629-637.
6. Kozlovskaya, L.I., Volok, V.P., Shtro, A.A. et al. (2021) *European Journal of Medicinal Chemistry*, 220, 113467.
7. Dueva, E.V., Tuchynskaya, K.K., Kozlovskaya, L.I. et al. (2020) *Antiviral Chemistry and Chemotherapy*, 28, 2040206620943462.
8. Zenchenko, A.A., Oslovsky, V.E., Varizhuk, I.V. et al. (2022) *Toxicology in Vitro*, 105355.
9. Zakharova M.Y., Kuznetsova, A.A., Uvarova, V.I. et al. (2021) *Frontiers in Pharmacology*, 12, 773198.

TARGETED PROTEOMICS FOR HEALTH ANALYTICS: OPPORTUNITIES AND CHALLENGES

E.A. Ponomarenko*[1]*, S.E. Novikova[1], E.V. Ilgisonis[1], O.I. Kiseleva[1], A.A. Kliuchnikova[1], S.N. Tarbeeva[1], M.A. Pyatnitskiy[1], A.V. Lisitsa[1], T.E. Farafonova[1], E.V. Poverennaya[1], A.T. Kopylov[1], Y.K. Paik[2], G. Hosseini Salekdeh[3], A. Urbani[4], V.G. Zgoda[1], A.I. Archakov[1]

[1]Institute of Biomedical Chemistry, Moscow, Russia
[2]Yonsei Proteome Research Center, Yonsei University, Seoul, Korea
[3]Department of Molecular Systems Biology, Cell Science Research Center, Royan Institute for Stem Cell Biology and Technology, ACECR, Tehran, Iran
[4]Area of Diagnostic Laboratories, Fondazione Policlinico Gemelli – IRCCS, Rome, Italy

The proteomic composition of a biological sample serves as the most important feature of a biological object, and it allows discriminating normal and pathological conditions. Targeted mass spectrometric analysis, namely, selected reaction monitoring (SRM) using synthetic isotopically-labeled internal standard (SIS), is the main alternative to the ELISA method for the analysis of diagnostically significant proteins.

The Russian Consortium is developing plasma analysis technology, which combines the chromosome-centric approach with bottom-up SRM SIS. This study was aimed to quantitatively analyze the proteins encoded by 643 genes of the four selected chromosomes during C-HPP (Chr18/Russia, Chr13/Korea, ChrY/Iran, ChrMt/Italy) in the blood plasma of healthy, clinically well-examined people (astronaut candidates) using SRM SIS technologies [1]. Fifty-four male subjects (age 20-47) were examined at the Institute of Medico-Biological Problems (Moscow, Russia).

The concentration for 205 proteins and also FDA-verified proteins [2] was accurately measured with SRM SIS assay, while quantitative proteomic profile of each sample was presented as personal QR-code. The concentration range covered by the SRM SIS technology was six orders of magnitude (from 10^{-6} to 10^{-11} M) in case of the analysis of whole plasma, and five orders of magnitude (from 10^{-7} to 10^{-11} M) in case of the analysis of depleted samples.

Proteins, which abundance levels are more or less stable in samples derived from healthy volunteers (inter-individual CV $\leq$40% and technical variability <20%), could be used as a pillar for creation SRM-assays for personal health analytics. There was no correlation between protein abundances and corresponding number of samples in which this protein was detected.

Based on the SRM SIS results, a prototype test system has been developed based on targeted mass spectrometric method for multiplex, quantitative analysis of in blood plasma. It seems that the developed prototype test system based on targeted mass spectrometric analysis of proteins can be considered as an alternative to methods using monoclonal antibodies.

The study was performed employing "Avogadro" large-scale research facilities, and was financially supported by the Ministry of Education and Science of the Russian Federation, Agreement No. 075-15-2021-993, unique project ID: RF00121X0004.

1. Kopylov, A.T., Ponomarenko, E.A., Ilgisonis, E.V. et al. (2019) *J. Proteome Res.*, 18(1), 120-129.
2. Novikova, S.E., Farafonova, T.E., Tikhonova, O.V. et al. (2020) *Biomed. Khim.*, 66(4), 294-316.

CHEMICAL READ-ACROSS PREDICTIONS OF ECOTOXICITY DATA

K. Roy

Drug Theoretics and Cheminformatics Laboratory, Department of Pharmaceutical Technology, Jadavpur University, Kolkata, India

The *in silico* modeling new approach methodologies (NAMs) represent a promising starting point among a wide range of alternative non-testing approaches for the safety evaluation of chemicals and filling the existing gaps in safety and ecosafety data. These computer-aided methods comprise structural alerts, grouping, read-across and quantitative structure–activity relationship (QSARs) approaches. As recently pointed out by the European Chemicals Agency and some authors, read-across is one of the most widely used alternative tools for hazard assessment, aimed at filling data gaps. Read-across not being a hardcore statistical approach, read-across based predictions appear to be more appropriate for small data sets. In the present talk, we discuss a new similarity based read-across algorithm for the prediction of toxicity (biological activity in general) of untested compounds from structural analogues [1]. Three similarity estimation techniques such as, Euclidean distance based similarity, Gaussian kernel function similarity, and Laplacian kernel function similarity are used in this algorithm. The quality of predictions depends on the selection of the distance threshold, similarity threshold, and the number of most similar training compounds. After toxicity prediction of test set compounds, the external validation metrics such as $Q^2_{ext\_F1}$, $Q^2_{ext\_F2}$, RMSEp were calculated. The computed metric values clearly justify the efficiency of the new read-across method and accuracy of the generated data by the proposed algorithm. A java based computer program (available at https://sites.google.com/jadavpuruniversity.in/dtc-lab-software/home) has also been developed based on the proposed algorithm which can effectively predict the toxicity of unknown chemicals after providing the structural information of chemical analogues. The tool (v4.0) is also able to handle graded data for classification-based predictions and generate different error and similarity-based measures which can be used for checking the reliability of quantitative predictions for new compounds. The new algorithm has been applied for predictions of different toxicity and ecotoxicity data sets. The new algorithm and the program can be used for the data gap filling, prioritizing existing and new chemicals, and for their risk assessments.

1. Chatterjee, M., Banerjee, A., De, P., Gajewicz-Skretna, A., Roy, K. (2022) *Environ. Sci. Nano*, 9, 189-203.

THE STATUS OF THE COMPUTER AIDED DRUG DESIGN: THEN, NOW AND FUTURE

G.N. Sastry

CSIR-North East Institute of Science and Technology, Jorhat, India

When humans are exposed to severe pain, injury or an unexplained weakness or lethargy to do things-this condition is identified with lack of health or the person is supposed to be inflicted with a disease. Thus, the health and disease are very closely connected and have an inverse relationship. Drug discovery has been a historic necessity. Natural choice is to look at natural sources such as plants, minerals, etc. Then, when people are able to classify different diseases, a systematic way of understanding the relationship between drug and disease was established and the therapeutics formulated. In this talk an attempt has been made to examine the role of different disciplines in drug design, and particularly based on in silico approaches. While the entry of powerful computers has undeniably altered the way in which drug design and discovery practiced in academia and industry, the evolution of these methods reveals the multifaceted nature of the perspective. Computer aided methods, were initially inspired by lock and key concept giving raised to structure and analog based approaches. However, deeper insights into the details of the receptors and metabolic pathways, ADMET reveal the limitations of one-drug one-receptor approach. Every therapeutic agent is essentially toxic and the dosage is a crucial factor to determine the optimal benefit as opposed to the severe side effects. Side effects of medicines warrant a much deeper understanding the pathophysiology of the disease as well as a fair idea on the therapeutic interventions of drugs in off-targets. Polypharmacology, network pharmacology, genomics and personalized medicine have become essential to satisfy the regulatory compliance and assess the risk and efficacy factors. In the last decade or so the amount of data available is so vast and it is often outside the limits of comprehension of rational thinking. The quest to convert the data into knowledge has necessitated adoption of data sciences approaches in drug discovery. While the modern medicines have provided a wealth of knowledge, great ability to treat disease and in general increase the life span of health populations, there are certain areas where the most modern approaches appear to fail. Interestingly the number of cases where the traditional wisdom which exist in the several parts of the world for thousands of years seem to provide solutions where modern medicine reaches a road block. Therefore, it is conceivable to think of a holistic approach integrating molecular modeling, mathematical modeling and data science approaches with a clear focus on providing right solutions to the needy patients. Therefore, future appear to lie in personalized medicine, where a large number of factors including the patient genetic, biochemical, anthropometric and life style parameters need to be considered before engaging in treatment. The scope and need to employ artificial intelligence and IoT approaches appeared to be indispensable not only in diagnostics but also in drug design.

NATURAL PRODUCTS DATABASES AS VALUABLE SOURCES OF BIOACTIVE STRUCTURES FOR VIRTUAL SCREENING

E.F. Albuquerque, L.F. Calado, C. Herrera-Acevedo, R.P.B. de Menezes, M.T. Scotti

Federal University of Paraíba, João Pessoa, Brazil

Computational approaches have played an increasingly prominent role in the natural product (NP)-based drug discovery. For example, the development and use of NP databases allow access to numerous chemical, biological, pharmacological, toxicological, and structural NP data. Currently, available NP databases, only a few the databases were sustainably managed and continually developed. Nevertheless, many NP databases are constantly growing and being updated and are regularly employed in drug design, e.g., ZINC, SuperNatural II. Free, open NP databases (e.g., the COlleCtion of Open Natural prodUcTs (COCONUT), comprising "flat" (lacking stereochemistry) NPs and stereochemically preserved NPs) have emerged as important online research tools, facilitating access to NP databases and addressing the limitations observed in many of these computational tools. Such NP databases allow for bulk downloads, enabling their use for in silico research, such as virtual screening. Other features, such as the geographical origin of secondary metabolites, have also become increasingly common in NP databases. SistematX (http://sistematx.ufpb.br) is a web portal of natural products developed at the Federal University of Paraíba, PB, Brazil, and originally introduced in 2018. It is an open-access database of secondary metabolites available to any research group.

SistematX[1] has emerged as a promising tool that connects many molecular properties to the reporting literature, facilitates the use and visualization of these properties, and provides information for chemosystematic studies, compound dereplication, and taxonomic correlations. These goals are achieved through the following features: (a) chemical retrieval by structure, simplified molecular-input line-entry system (SMILES) code, compound name, and plant species; (b) inclusion of chemical structures and characteristics important for NP chemistry in search results; and (c) storage of search results according to the best practices in the field, including curated chemical structures, taxonomy of the plant from which the compound was isolated, bibliographic reference(s), and Global Positioning System coordinates. Future directions in NP databases emphasize data curation, particularly dereplication, to minimize the probability of rediscovering known compounds.

Structures in SistematX are curated in two steps. A designated SistematX administrator first reviews the structure and data in the original source paper, book, etc.) prior to registration in the database. The 2D and 3D chemical structures are then standardized using a ChemAxon JChem Standardizer,10,11 which validates all structural features, including stereochemistry. The 3D structures obtained using JChem Standardizer are considered reliable because they are canonized (i.e., hydrogens are added, the aromatic form is converted, and the molecular graphs are cleaned in three dimensions). Additionally, this tool generates and optimizes conformers of the initial structure. Any molecules still presenting structural problems after these processes are manually corrected using ChemAxon MarvinSketch before registration in SistematX. Since 2018, SistematX has reported 5.126 new users and greater than 8950 sessions, with approximately 40% located outside of Brazil. SistematX currently has 9514 unique secondary metabolites arising from 20 934 botanical occurrences across 5 families: Asteraceae: 2574 (7879 occurrences), Apocynaceae: 372 (620 occurrences), Annonaceae: 1898 (5335 occurrences), Lamiaceae: 4097 (6357 occurrences), and Solanaceae: 573 (743 occurrences). The increased number of secondary metabolites registered in SistematX has allowed the software developer group to prepare virtual screening studies that have identified new structures with potential activity against several diseases (e.g., Chagas diseases, Leishmaniasis, and SARS-CoV-2) and chemotaxonomic studies using self-organizing map (SOM) algorithms.

1. Costa, P.O.R., Lucena, L.F., Silva L.M.A., Zocolo, G.J. et al, (2021) *Journal of Chemical Information and Modeling*, 61, 2516.

COMPUTATIONAL STUDIES ON GREEN PESTICIDES

N. Khazanov, S. Kozokaro, L. Iktelat, O. Kaspi, P. Clarke, H. Senderowitz

Department of Chemistry, Bar Ilan University, Ramat Gan, Israel

Population growth coupled with reduction in agriculture land compromise food security for millions of people worldwide. To increase the yields of agriculture products in field and storage, pesticides are often used. However, the massive usage of many currently available pesticides leads to pollution, residual toxicity and adverse effects on human health. Thus new, so-called "green pesticides" are critically needed. The development of such pesticides requires, in turn, the identification of relevant targets in plants pathogens and the development of compounds inhibiting them.

In this work, we describe the computer assisted modeling and discovery of compounds targeting different pathogenic proteins including enzymes belonging to the bacterial quorum sensing (QS) machinery, ABC exporters and other efflux pumps and enzymes that participate in the construction of cell walls of oomycetes. QS is a population density-dependent regulatory mechanism used by bacteria to coordinate virulence. Depending on the specific bacteria, the QS machinery is comprised of protein(s) that synthesize signaling molecule(s) and protein(s) that receive them. Upon binding the signal molecule(s) the receiving protein(s) initiate a cascade of actions that leads to the synthesis of various virulence factors. Thus, inhibiting either the synthesizing or the receiving proteins is expected to reduce virulence, importantly without killing the bacteria. Efflux pumps control the export of chemicals from within bacterial cells. Some of the exported chemicals are QS inhibitors. Thus inhibition of pumps either alone or in combination with QS enzymes, is another viable strategy to combat virulence.

For QS proteins and ABC efflux pumps, we have modeled the relevant protein structures, identified putative binding sites, performed extensive VS of ~30M compounds per target using a combination of ligand-based and structure-based techniques and identified multiple virtual hits. Working in collaboration with the group of Prof. Yedidia from the Volcani center, the resulting hits were evaluated for their biological activity. Some of the identified compounds were indeed found to effectively block bacterial virulence by directly inhibiting the AHL synthase (ExpI) and/or the AcrAB-TolC efflux pump of *P. brasiliense*, a plant pathogen that is responsible for soft rot disease in a wide range of ornamental and vegetable crops.

Enzymes that participate in the construction of cell walls of oomycetes could be inhibited by several types of compounds including linear aptamers (small peptides). Using a yeast two-hybrid approach, the groups of Profs. Pesaresi and Masiero from the University of Milano, have indeed identified sets of aptamers inhibiting various cell-wall constructing enzymes. In this work we also present a chemoinformatic workflow for the derivation of predictive QSAR models for these aptamers. Typically, classification models require sufficiently large and balanced (i.e., with similar numbers of samples for each class) datasets. However, neither of these requirements are met in the present case. Therefore, we introduce a workflow with novel features such as usage of sequence-aware global molecular descriptors, and a statistics-based module deigned to accurately generate synthetic data, thus meeting both requirements. We applied the new algorithm to five aptamer datasets and demonstrated that in many cases, datasets augmented by the synthetic data gave rise to models with better performances than models derived from the original dataset. Moreover, such models performed better than models derived from scrambled data.

1. Pun, M. et al. (2021) *Frontiers in Plant Sciences*, 12, 671807.
2. Joshi, J.R. et al. (2021) *Annual Review of Phytopathology*, 59, 153-190.
3. Joshi, J.R et al. (2020) *ACS Chemical Biology*, 15(7), 1883–1891.
4. Joshi, J.R et al. (2016) *Scientific Reports*, 6, 38126.

MOLECULAR DYNAMICS STUDIES ON THE INTERACTIONS BETWEEN SARS-COV-2 SPIKE PROTEIN AND HACE2 OR MABS

W. Zhu

Drug Discovery and Design Center, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai, China

The spike protein of SARS-CoV-2 (CoV-2-S) mediates the virus entry into human cell, making it a suitable target for drug discovery and vaccine development. Experimental studies showed stronger binding affinity of the spike RBD (receptor binding domain) of CoV-2-S to human ACE2 than SARS-CoV spike (CoV-S), but similar binding affinity of CoV-2 entire spike to CoV-S's. Furthermore, the emergence of various SARS-CoV-2 mutations, e.g., E484K, Delta and Omicron variants, show significant changes in binidng ACE2 and obvious resistance to the neutralization of vaccine sera, leading to the risk of ineffectiveness of currently available drugs and mAbs.

To explore the binding mechanism, we calculated the binding affinities of the RBDs to ACE2 and simulated the transitions between ACE2-inaccessible and -accessible conformations of the spike. We found that although the ACE2-binding strength of CoV-2-S RBD is much stronger than CoV-S RBD, CoV-2-S has much less ACE2-accessible conformation, making the binding affinity of entire protein decreased in comparison with that of CoV-2-S RBD. Further analysis revealed key interactional residues for strong binding and five potential ligand-binding pockets for drug research.

To explore the potential risk of immune evasion from various mutations, we calculated the binding affinities of 26 antibodies to wild type (WT) spike protein and to the protein harboring E484K mutation, respectively. The results showed that most antibodies (~85%) have weaker binding affinities to the E484K mutated spike protein than to the WT, indicating the high risk of immune evasion of the mutated virus from most of current antibodies. Based on binding free energy decomposition, we predicted that the mutation of 4 more residues on the RBD of spike protein, *viz*., F490, V483, G485 and S494, may have high risk of immune evasion, which we should pay close attention on during the development of new mAb therapeutics.

In explore the changes of the binding affinity between WT/Delta/Omicron RBDs and ACE2/mAbs, we performed MD simulations for the complexes formed by the WT, Delta and Omicron variant RBDs and hACE2 (600 ns in total for each system), respectively, and calculated the binding free energy (ΔG) by MM/GBSA. It was found that the binding affinity of Omicron RBD to ACE2 is much weaker than that of Delta RBD, but comparable binding affinity to WT RBD. Therefore, the Omicron variant might have milder infectiveness than Delta, if the infection capability is mainly determined by RBD-ACE2 binding affinity. We also performed ELISA bioassay, which suggest that Omicron variant possesses comparable binding affinity to human ACE2 in comparison with the wild type SARS-CoV-2, but much weaker binding affinity than Delta variant. In addition, the MD simulations indicate that the Omicron has high risk of immune evasion. Accordingly, close attention should be paid intensively to Omicron as its high immune evasion risk enables its easy transmission.

1. Peng, C., Zhu, Z., Shi, Y. et al. (2020) *J. Phys. Chem. Lett.*, 11, 10482.
2. Wu, L., Peng, C., Yang, Y. et al. (2021) *Briefings in Bioinformatics*, 23(1) bbab383.
3. Wu, L., Zhou, L., Mo, M. et al. (2022) *Signal Transduction and Targeted Therapy*, 7, 8.

ORAL PRESENTATIONS

DATABASE OF ANTIMICROBIAL ACTIVITY AND STRUCTURE OF PEPTIDES (DBAASP) – FINDING A WAY OUT OF MICROBIAL RESISTANCE

***E. Alimbarashvili*, N. Samsonidze, M. Grigolava, M. Pirtskhalava**

Ivane Beritashvili Center of Experimental Biomedicine, Tbilisi, Georgia

Microbial resistance towards conventional antibiotics has become a global problem threatening the public healthcare system. Owing to this problem, discovering new therapeutic drugs is of utmost importance. Antimicrobial peptides (AMPs) are a diverse group of naturally occurring molecules that play a key role in the innate immune system. Microorganisms are having more difficulties developing resistance towards AMPs than towards antibiotics – indicating mainly different mechanisms of action of these two groups of antimicrobials. AMPs are considered promising anti-infectives with the potential to tackle the aforementioned issue of microbial resistance. The Database of Antimicrobial Activity and Structure of Peptides (DBAASP) was launched in 2014 and serves as the repository for comprehensive information on both natural and synthetic AMPs, including experimental data, spatial structures, MD models, and corresponding linkages to the other databases [1]. The number of peptides stored in DBAASP has already exceeded 19000, among which are antibacterial, antiviral, anticancer, antiparasitic peptides with the data on their cytotoxic/hemolytic activities. The infrastructure of DBAASP allows users to obtain desired data conveniently. Therefore, DBAASP represents a valuable source for the scientific community to address the problem of microbial resistance development. Its data and prediction tools are widely used to design new antimicrobials prior to their actual synthesis, thus decreasing production costs.

In order to use AMPs effectively, comprehensive knowledge of their functioning in living organisms is required as they are naturally produced molecules. The innate immune system of animals, plants, and other organisms uses the rationally developed cocktail of AMPs while interacting with pathogenic or non-pathogenic microbes. Synergism is a widespread event among AMPs. The variety and synergism are the features of defense peptides responsible for optimizing defense expenses and combatting microbial resistance. Due to the increasing interest in the synergistic potential of AMPs, the information about synergistic interactions between AMPs, between AMP and antibiotics/metal ions appears faster than ever before. Consequently, DBAASP is developing to become an exhaustive repository for the data on synergy.

The current version of DBAASP stores information about more than 3000 synergistic interactions of AMPs, antibiotics, and metal ions against microbial strains. To the best of our knowledge, presently, no other database provides thorough information on the synergistic activities of AMPs. Accordingly, DBAASP will grow into a desirable resource to investigate synergistic interactions of antimicrobials, which may lead to discovering new mechanisms of action, concentrations, and combinations of peptide-based drugs. It will help remove the hindrances related to the use of AMPs and improve our understanding regarding drug-microbe interactions. Moreover, it might become the solution to combat multidrug-resistant (MDR) pathogens – a remaining issue and a way to use antibiotics towards which target pathogens have already become unresponsive.

This work was supported by Shota Rustaveli National Science Foundation of Georgia (SRNSFG) [grant No. FR-21-3368].

1. Pirtskhalava, M., Amstrong, A.A., Grigolava, M., Chubinidze, M., Alimbarashvili, E., Vishnepolsky, B., Gabrielian, A., Rosenthal, A., Hurt, D.E., Tartakovsky, M. (2021) *Nucleic Acids Research*, 49(D1), D288–D297.

A STRUCTURAL/DYNAMIC MODEL OF SARS-COV-2 SPIKE TRANSMEMBRANE DOMAIN IN CONJUNCTION WITH THE HR2 REGION: IMPLICATIONS FOR MEMBRANE FUSION

***E.T. Aliper*[1], *N.A. Krylov*[1,2], *D.E. Nolde*[1,2], *A.A. Polyansky*[1,3], *R.G. Efremov*[1,2,4]**

[1]*Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Moscow, Russia*
[2]*National Research University Higher School of Economics, Moscow, Russia*
[3]*Department of Structural and Computational Biology, Max Perutz Labs, Vienna, Austria*
[4]*Moscow Institute of Physics and Technology (State University), Dolgoprudny, Russia*

The spike (S) protein of SARS-CoV-2 effectuates membrane fusion and virus entry into target cells. Its transmembrane domain (TMD) is a homotrimer of α-helices anchoring the spike in the viral envelope. Although S-protein models available to date include TMD, its precise configuration has only been given brief consideration. Understanding viral fusion entails realistic TMD models, while no reliable approaches towards predicting the 3D structure of transmembrane (TM) trimers exist. Immediately upstream of the TMD is a domain known as HR2. Apart from being involved in the complex refolding of the spike, HR2 is believed to interact with the viral envelope, as it is brought into contact with the target membrane, which has been observed for HIV's protein gp41 from the same family of fusion proteins. Here, we adopted diverse computational tools to model the spike TMD (S-TMD) based solely on its primary structure and used the resulting model to study a larger element of the fusion machinery, TMD linked to HR2, i.e., an anchor plus a "membrane-fixing factor".

We performed amino acid sequence pattern matching and compared molecular hydrophobicity potential (MHP) distribution on the helix surface against TM homotrimers with known 3D structures. Eventually, the TMD of the tumour necrosis factor receptor 1 (TNFR-1) was selected for template-based modelling. Adjusting so-called "dynamic MHP portraits" and residue variability motifs, we iteratively built an all-atom homotrimer model of the S-TMD, whereof each helix possessed two overlapping interfaces interacting with either of the remaining helices. The interfaces included conservative residues like I1216, F1220 and M1229. The stability of this model was tested in all-atom molecular dynamics (MD) simulations in a POPC bilayer mimicking the viral envelope and compared to several alternative configurations, including a recent NMR structure of a trimerised peptide, several other available models and a set of mutant forms.

Unlike other configurations, our model trimer remained extraordinarily tightly packed over a microsecond-range MD and retained its stability when palmitoyl chains were added at cysteine residues located downstream in accordance with experimental data. Overall, the resulting model of S-TMD conforms to the known basic principles of TM helix packing.

S-TMD was then linked to the HR2 region, modelled based on an NMR structure of the same domain in SARS-CoV's spike. Models of monomeric and trimeric HR2 both on its own and linked to the TMD were subjected to MD simulations to evaluate the behavior and mutual influence of the two domains when part of one molecule, possibly relevant to understanding viral fusion. Trimeric HR2 remained intact in water, and so did trimeric HR2 linked to TMD in a model POPC bilayer: throughout the MD simulation, both domains remained stable. Previously SARS-CoV-2's spike HR2 had been experimentally demonstrated to have limited affinity for lipid bilayers, and our MD simulations agree with these observations: on its own, monomeric HR2 did not remain membrane-bound. Linking it to a monomer of TMD, on the other hand, resulted in more frequent contacts with the bilayer. The TM anchor thus significantly increases the ability of HR2 to be adsorbed on a model membrane mimicking the viral envelope, although such interaction is still too weak to consider HR2 as a peripheral membrane domain. We hypothesise that other factors than the TMD come into play for proper binding of HR2 to the membrane to take place during the early stages of fusion.

This work was supported by RSF grant No. 18-14-00375.

INTERACTION AND INHIBITION OF ALPHA-GLUCOSIDASE WITH SELECTED MONOTERPENES

P. Behera, A. Mukerjee, S. Balaji

Department of Biotechnology, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, India

Diabetes mellites is a global health problem, and different methods are being adopted to either eradicate or control this disease. One of the methods adopted is the usage of α-glucosidase inhibitors. The inhibitory effects of many monoterpenes for instance, carvacrol, carveol, camphene, α-pinene, p-cymene, limonene, citronellal, citronellol and 1,8-cineole were analyzed for their inhibitory potential against α-glucosidase enzyme (maltase). The quantitative structure-activity relationships of monoterpenes and their derivatives were studied, for instance carvacrol contains hydroxyl group in contrast to cymene had a significant result, whereas the aliphatic side chains of carvacrol have less effect on the inhibitory activity. The molecular interactions that are due to the various functional groups were studied. The stability of the ligand-protein complex was revealed by the low binding energy (kcal/mol). The results indicated that these inhibitors have the potential to become an anti-diabetic drug for the treatment of Diabetes mellites.

AN INSIGHT INTO THE ORIGIN OF MICROTUBULE-CURLING EFFECT OF PODOPHYLLOTOXIN ESTERS: MOLECULAR DYNAMICS STUDY

***A. Borovik*[1], *N. Zefirov*[1] , *E. Radchenko*[1], *S. Kuznetsov*[2], *O. Zefirova*[1]**

[1]*Lomonosov Moscow State University, Russia*
[2]*Institute of Biological Sciences, University of Rostock, Germany*

Natural lignan podophyllotoxin (PF) possesses anticancer activity due to its ability to inhibit polymerization of α,β-tubulin to microtubules (MTs). The compound is not applied in chemotherapy because of its high general toxicity and numerous PF analogues were synthesized in a search of more safe drugs. In the course of these works a series of bridged C^4-esters were obtained and showed cytotoxicity in a wide range from subnanomolar to micromolar [1–4]. Tested at a single concentration of 10 mkM the compounds were found either to induce MTs shortening or depolymerization, typical for podophyllotoxin and its active derivatives, or to alter the dynamics of microtubule cytoskeleton in unusual manner by stimulating the formation of involuted structures defined as "curled MTs". To the best of our knowledge this action was described only for C^4-PF esters with bridged residues. Therefore, in the present work we additionally synthesized five novel C^4-PF esters with alicyclic moieties, tested them by immunofluorescence microscopy at different concentrations and performed molecular dynamics study with the purpose to hypothesize about the origin of MTs curling effect in cancer cells.

The compounds were obtained by Steglich esterification of podophyllotoxin with commercially available alicyclic acids. Immunofluorescent microscopy of cancer cells A549 treated with novel esters at different concentrations gave evidence that the "curling" of microtubules takes place at one of the first steps of their depolymerization. Considering that at substoichiometric concentrations podophyllotoxin is able to suppress the dynamic instability of MTs due to the weak association of free tubulin–PF complex with shortened MTs, it was hypothesized, that free tubulin complexes with bridged podophyllotoxin esters may have noticeable structural differences that enable them to copolymerize with shortened MTs leading to "MTs" with different dynamics and “curled” morphology. To check if this is possible in principle, the molecular dynamics simulations of the behavior of tubulin and two complexes *ligand – tubulin* (PDB ID: 1SA1) were performed in the CHARMM36 / CGenFF 4.4 force field using GROMACS 2020.3 program [5, 6].

The molecular dynamics study revealed notable differences in the curved conformations (important for the MTs depolymerization caused by the ligands of colchicine binding site) of tubulin dimer in a complex with podophyllotoxin and its ester with 1-adamantaneacetic acid with very pronounced MTs-curling effect. Earlier based on the molecular docking data, we hypothesized that MTs curling may be a consequence of proximity the bridged moieties to GTP molecule, tightly bound to α-subunit of tubulin dimer [4]. New data allow to hypothesize, that the "curling" of microtubules caused by bridged podophyllotoxin C^4-esters occurs due to a change in the curvature of curved conformation of tubulin dimer.

The research was carried out in the framework of the Russian state assignment No. 121021000105-7.

1. Zefirov, N.A., Kruth, A., Wobith, B. et al. (2018) *Mendeleev Commun*., 28, 47.
2. Zefirov, N.A., Evteeva, Yu.A., Wobith, B. et al. (2019) *Struct. Chem*., 30, 465.
3. López-Pérez, J.L., del Olmo, E., de Pascual-Teresa, B. et al. (2004) *Bioorg. Med. Chem*., 14, 1283.
4. Zefirov, N.A., Lavrushkina, E.A., Kuznetsov, S.A., Zefirova, O.N. (2019) *Biomeditsinskaya Khimiya (in Russ.)*, 65, 86.
5. Huang, J., MacKerell, A.D.J. (2013) *J. Comput. Chem*., 34, 2135.
6. Lee, J., Cheng, X., Swails, J.M. et al. (2016) *J. Chem. Theory Comput*., 12, 405.

QSPR ANALYSIS IN PHOTONICS

A.A. Buglak[1,2]

[1]*Saint Petersburg State University, Saint Petersburg, Russia*
[2]*Kazan Federal University, Kazan, Russia*

The objective of our study is to analyze available Quantitative Structure-Property Relationship (QSPR) models allowing to predict the photodynamic activity of different materials and molecules demanded by biomedicine. QSPR can be applied to the analysis of absorption spectra, fluorescence intensity and wavelength, photolysis quantum yield, triplet state formation and $^1O_2$ generation, as well as phototoxicity. Models predicting absorption wavelength and oscillator strength are probably the most developed ones and are widely used for the analysis of a wide range of organic chromophores since early 2000s. Prediction of fluorescence intensity and wavelength is used for different types of fluorophores. QSAR models predicting phototoxicity stand along as the models developed on living organisms: usually cancer cells or microbiological culture. Recently, we have established QSPR models for singlet oxygen generation in O_2 rich solutions by organic and metal-organic photosensitizers.

First, we analyzed the quantum yields of singlet oxygen generation (Φ_Δ) by 29 pteridines, namely pterins and flavins, in aqueous solutions. We found that Φ_Δ of pterins depends on the ionization potential, electronegativity and some minor descriptors. The multiple linear regression (MLR) method was used to build a QSPR model allowing to predict Φ_Δ with "leave-one-out" cross-validation q^2 equal 0.881 and predictive R^2 equal 0.873.

The quantum yields of triplet state generation (Φ_T) of 26 furocoumarins (psoralens and angelicins) were found to be significantly correlated with T_1 state energy and Jhetv descriptor ($R=0.792$, $q^2=0.865$, predictive $R^2=0.897$).

The random forest (RF) QSPR model for a dataset containing 32 porphyrins, chlorins, bacteriochlorins, and metalloporphyrins demonstrated high R (0.974) and predicting ability towards the test set (0.875).

For a dataset containing >75 BODIPYs in toluene, tetrahydrofuran and acetonitrile MLR models performed significantly better than other machine learning methods (RF, SVM) and showed sufficient statistical parameters ($R=0.88$-0.91, $q^2=0.62$-0.69).

As a result, we proved that QSPR and machine learning techniques is useful for the prediction of Φ_Δ values in different media and virtual screening of new fluorophores with improved photosensitizing ability. Our approach seems to be more fundamental than a trivial "structure – phototoxicity" relationships. We believe that our QSPR results can be useful not only to create new photosensitizers for antimicrobial and antitumor photodynamic therapy, but also can be exploited in drug design, blood sterilization, wastewater treatment, sunlight-activated herbicides and insecticides.

PHF10 – THE SUBUNIT OF PBAF CHROMATIN REMODELING COMPLEX: STRUCTURE AND FUNCTION PREDICTIONS

***A. Chugunov*[1], *N. Potapova*[2], *N. Soshnikova*[3]**

[1]*Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Moscow, Russia*
[2]*Institute for Information Transmission Problems (Kharkevich Institute), Moscow, Russia*
[3]*Engelhardt Institute of Molecular Biology, Moscow, Russia*

Transcription activation factors and multisubunit coactivator complexes get recruited at specific chromatin sites *via* protein domains that recognize histone modifications. Single PHDs (plant homeodomains) interact with differentially modified H3 histone tails. Double PHD finger (DPF) domains possess a unique structure different from PHD and are found in six proteins: histone acetyltransferases MOZ and MORF; subunits of chromatin remodeling complexes BAF (DPF1–3) and PBAF (PHF10). Among them, PHF10 stands out due to the DPF sequence, structure, and functions. PHF10 is ubiquitously expressed in developing and adult organisms as four isoforms differing in structure (the presence or absence of DPF) and transcription regulation functions.

Despite the importance of the DPF domain of PHF10 for transcription activation, its structure remains undetermined. We performed homology modeling of the human PHF10 DPF domain [1] and determined common and distinct features in structure and histone modifications recognition capabilities, which can affect PBAF complex chromatin recruitment. We also traced the evolution of DPF1–3 and PHF10 genes from unicellular to vertebrate organisms. The data reviewed suggest that the DPF domain of PHF10 plays an important role in SWI/SNF-dependent chromatin remodeling during transcription activation.

This research was funded by the Russian Science Foundation, grant No. 21-14-00258 (to S.N.).

1. Chugunov, A.O., Potapova, N.A., Klimenko, N.S., Tatarskiy, V.V., Georgieva, S.G., Soshnikova, N.V. (2021) *Int. J. Mol. Sci.* 22(20), 11134.

HARNESSING MACHINE LEARNING FOR DRUG DISCOVERY

G. Bitencourt-Ferreira, W.F. de Azevedo, Jr.

School of Health and Life Sciences. Pontifical Catholic University of Rio Grande do Sul (PUCRS), Porto Alegre, Brazil

Docking simulation allows us to predict the orientation of ligands into a protein-binding pocket. Even with the continuing progress of docking and machine learning programs, we still miss a fully automated tool to integrate the most recent docking and machine learning techniques. This integration would open the possibility of developing a model targeted to a protein of interest, exploring the scoring function space [1]. That is the motivation behind SAnDReS 2.0; integration in one tool of recent programs and databases for docking, machine learning, structural and binding affinity data (PDBbind) [2].

SAnDReS 2.0 is an open-source suite of programs (GNU License) to develop machine learning models to predict binding affinity based on crystal structures or poses. It has internal ligand datasets of binding affinity derived from PDBbind (2020 version) and automatically combines binding with structural data. These datasets have crystal structures only since over 98 % of PDB entries with binding affinity data (IC_{50}, K_d, and K_i) use the X-ray crystallography technique [1].

Our approach brings together recent tools for docking and machine learning modeling. We have the newest version of AutoDock Vina (version 1.2.3) and MGLTools1.5.7 [3] for docking and file preparation. Also, we employ the latest version of Scikit-Learn (version 1.0.2) (April 2022). We have sixty-four regression methods allowing us to explore the scoring function space with efficiency. Half of these methods have cross-validation. SAnDReS 1.0 had nine regression methods and employed three features to build a mixed polynomial equation [4]. We do not have these limitations in SAnDReS 2.0.

SAnDReS 2.0 has two approaches to docking. The first employs three ligand-centering schemes: center of mass (CM), electric center (EC), and geometric center (GC). The CM is the standard centering adopted by Vina. These additional ligand-centering schemes are new in SAnDReS 2.0 and improve docking accuracy. In this first approach, we run three dockings for each entry in the dataset and select the best protocol based on the docking root-mean-square deviation (RMSD). The second approach runs dockings with CM only. Additionally, it is possible creating machine learning models using the results from external docking programs (e.g., Molegro Virtual Docker (MVD)) [5].

To our knowledge, SAnDReS is the first computational tool to integrate the machine learning methods of Scikit-Learn 1.0.2 and the energy terms of Vina (version 1.2.3) and MVD to explore the scoring function space.

This study was financed by the CAPES (Finance Code 001), PRO-Stricto Scholarship Program – PUCRS, and CNPq (309029/2018-0).

1. Heck, G.S. et al. (2017) *Curr. Med. Chem.*, 24(23), 2459–2470.
2. Wang, R. et al. (2004) *J. Med. Chem.*, 47(12), 2977–2980.
3. Eberhardt, J. et al. (2021) *J. Chem. Inf. Model.*, 61(8), 3891–3898.
4. Xavier, M.M. et al. (2016) *Comb. Chem. High Throughput Screen.*, 19(10), 801–812.
5. Thomsen, R., Christensen, M.H. (2006) *J. Med. Chem.*, 49(11), 3315–3321.

AMYLOIDOGENIC PEPTIDES: NEW CLASS OF ANTIMICROBIAL PEPTIDES WITH THE NOVEL MECHANISM OF ACTIVITY

***O.V. Galzitskaya*[1,2], *S.Y. Grishin*[1], *S.R. Kurpe*[1], *A.V. Panfilov*[1], *A.V. Glyakina*[1,3], *A.P. Kochetov*[4], *S.V. Kravchenko*[5], *P.A. Domnin*[6,7], *A.K. Surin*[1,4], *V.N. Azev*[4], *S.A. Ermolaeva*[7]**

[1]*Institute of Protein Research, Russian Academy of Sciences, Pushchino, Russia*
[2]*Institute of Theoretical and Experimental Biophysics, Pushchino, Russia*
[3] *Institute of Mathematical Problems of Biology and Keldysh Institute of Applied Mathematics, Pushchino, Russia*
[4]*The Branch of the Institute of Bioorganic Chemistry, Pushchino, Russia*
[5]*Institute of Environmental and Agricultural Biology (X-BIO), Tyumen State University, Tyumen, Russia*
[6]*Biology Faculty, Lomonosov Moscow State University, Moscow, Russia*
[7]*Gamaleya Research Centre of Epidemiology and Microbiology, Moscow, Russia*

The development and testing of new antimicrobial peptides (AMPs) is due to the high resistance of pathogenic bacteria to traditional antibiotics. The issue is particularly acute for the development of new antimicrobials capable of inhibiting the growth of multidrug resistant pathogens such as *Pseudomonas aeruginosa* and *Staphylococcus aureus*. Most AMPs achieve these goals through mechanisms that disrupt the normal permeability of the cell membrane, ultimately leading to the death of the pathogenic cell. We have developed a unique combination of membrane penetrating peptide and amyloid prone peptides to create hybrid peptides. While the cell-penetrating peptide allows the peptide to enter the bacterial cell, the amyloidogenic region provides an antimicrobial effect by co-aggregating with functional bacterial proteins. We evaluated the antimicrobial effects of six synthesized hybrid peptides, which were obtained based on the sequences of the S1 ribosomal protein of *P. aeruginosa* and *S. aureus*. It is important that some peptides demonstrated high antimicrobial activity comparable to the antibiotic gentamicin sulfate against pathogenic strains of methicillin-resistant *S. aureus* (MRSA), *S. aureus*, and *P. aeruginosa* [1,2]. These peptides showed no toxicity to eukaryotic cells. Our study demonstrates the promise of hybrid peptides based on the amyloidogenic regions of the S1 ribosomal protein for the development of new antimicrobials against Gram-positive and Gram-negative bacteria resistant to traditional antibiotic.

This study was supported by the Russian Science Foundation, grant No. 18-14-00321.

1. Grishin, S.Y., Domnin, P.A., Kravchenko, S.V. et al. (2021) *Int. J. Mol. Sci.*, 22, 9776.
2. Kravchenko, S.V., Domnin, P.A., Grishin, S.Y. et al. (2022) *Int. J. Mol. Sci.*, 23, 524.

SYSTEM PHARMACOLOGY IN DRUG DISCOVERY

S. Gharaghani

Laboratory of Bioinformatics and Drug Design (LBD), Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran

The "one-disease, one-target, one-drug" strategy has been the standard method for discovering new drugs in pharmaceutical research for fifty years. Therefore, methods for discovering new drug candidates use a library of compounds, one target, and an assay to measure the biological activity of these compounds against a single target protein. Traditional computational drug discovery methods such as molecular docking, pharmacophore, and quantitative structure-activity relationship (QSAR) models have been carried out to design potent compounds against the selected target. Unfortunately, QSAR as a ligand-based model cannot consider the structural information of the target and, in many studies, cannot predict selective drugs. Therefore a meaningful decrease in the discovery rate and design of selective and potent drug candidates has been observed. In this regard, alternative approaches are needed to increase our knowledge about the mechanisms of disease and the effect of a drug that subsequently leads to tremendous success for pharmaceutical research. Recent advances in systems biology have shown that most drugs interact with multiple targets, and considering one target in drug discovery projects is not rational. Advances in genome-scale metabolic models, structural bioinformatics, and systems pharmacology have caused the identification of new drug targets at the system level. Therefore in this lecture, I will describe the current state of the systems pharmacology in drug discovery and the results of my research in this area.

USE OF FACTORIAL DESIGN TO OPTIMIZE IMIDAZO[1,2-a]AZINES AS INHIBITORS OF COX ISOFORMS

M.E. Campos-Aldrete, M. Guerrero-González

Organic Chemistry Department, Escuela Nacional de Ciencias Biológicas-Instituto Politécnico Nacional, Mexico City, Mexico

Imidazo[1,2-a]azines with the acid group decreased the inflammatory process in murine models and the effect was attributed to the inhibition of cyclooxygenase (COX) enzymes, well known as prostaglandin-endoperoxide synthase (PTGS).

The optimization of a series of imidazo[1,2-a]azines was performed using the reduced factorial design 2^{3-1}. The inhibitory effect of 5 acid derivatives of imidazo[1,2-a]azines and standards ibuprofen and indomethacin were evaluated *in vitro* on the enzymes COX isoforms, it was observed that the different substituents provide different inhibition profiles, highlighting that the group of imidazo[1,2-a]pyridines is more active than the bioisosteric imidazo[1,2-a]pyrimidines, these results were analyzed *in silico* Docking to recognize the structural elements necessary for the inhibition of the targets. Conclusions: The IC_{50} to COX1 (µM) was obtained where the compounds showed 4a (2.72), 4b (3.94), 5a (7.29), 5b (63.26), 6a (12.93) indomethacin (0.13) ibuprofen (0.2) and IC50 in COX_2 (µM) 4a (1.89), 4b (2.39), 5a (8.08), 5b (41.15), 6a (5.86) indomethacin (0.09) ibuprofen (0.125), through factorial design it was possible to optimize the inhibitory response on therapeutic targets obtaining molecule 4a as a result of factorial analysis[1].

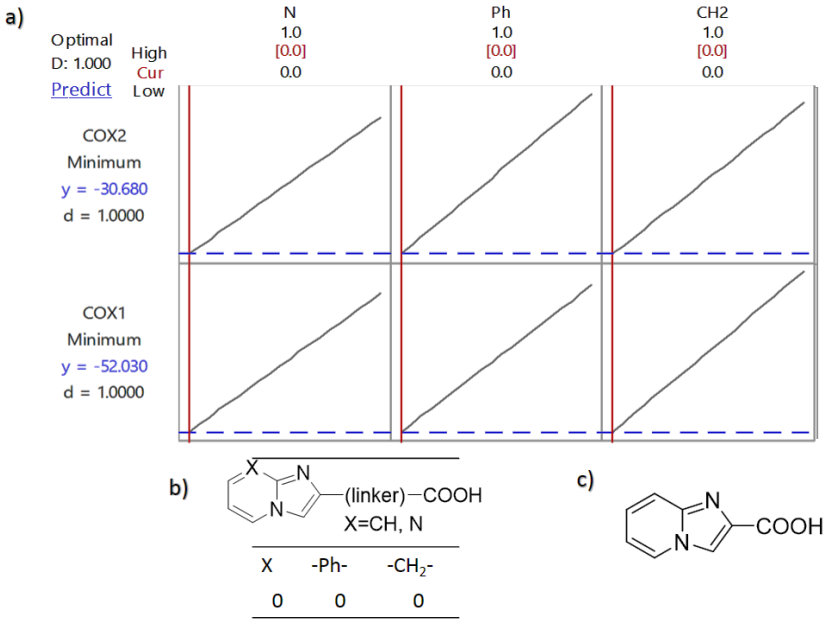


Figure 1. Optimization by the reduced factorial design 2^{3-1}. a) graphic optimization results in minitab 17. b) analysis of the results where it is observed the elements that the molecule must have that can improve the IC_{50} concentration to inhibit the 2 isoforms of COX. c) optimized inhibitor structure to carry it to synthesis and evaluation.

1. Martins Fukuda, I., Fidelis Pinto, C.F., dos Santos Moreira, C., Morais Saviano, A., Rebello Lourenço, F. (2018) *Braz. J. Pharm. Sci.*, 54(Special), e01006

DEVELOPMENT OF THE "VSAfiR" METHOD AND ITS APPLICATION IN THE DEVELOPMENT OF ANTIEPILEPTICS

M. E. Campos-Aldrete, M. Guerrero-González

Organic Chemistry Department, Escuela Nacional de Ciencias Biológicas-Instituto Politécnico Nacional, Mexico City, Mexico

Epilepsy is the most common severe non-infectious chronic neurological condition in the world. Only between 60-70% of patients respond to current drugs. According to the above, there is a need to develop new antiepileptic molecules.

Zolpidem is among the molecules with antiepileptic potential [1], however there is little information on its antiepileptic activity, and it is also necessary to optimize its antiepileptic activity. Achieving the optimization of the pharmacological activity of any therapeutic entity is a task that consumes a considerable amount of time and resources, which ends up translating into millions of dollars and years to achieve this goal.

Our work aims to introduce an algorithm of virtual screening methodology for save resources in molecular optimization programs, which we call "VSAfiR" (Virtual Structure Affinity Relationship), what we tested in the developing anti-epileptic molecules taking Zolpidem as a reference drug, using only data from molecular docking experiments, we were able to obtain structural changes that allow rational optimization of antiepileptic activity.

To develop the algorithm, 98 molecules were first proposed, which were taken to docking experiments where we evaluated the best poses in the benzodiazepine binding site in $GABA_A$ receptor, from these poses the affinity energy for each structural change was taken and modeled using the VSAfiR method. to select the functional groups that would enhance the antiepileptic response. Selected the substituents that *in silico* improved the binding to the receptor, we proceeded to the synthesis and characterization by NMR spectroscopy to confirm the chemical structure of each molecule. Once the molecules were prepared, we proceeded to determine the LD_{50} (lethal dose 50) and the evaluation of the antiepileptic activity how ED_{50} (effective dose 50), as well as the myorelaxant and sedative effect in murine models using methods that allow the use of a minimum number of animals.

The developed algorithm allowed selecting structural changes that allowed the development of antiepileptic molecules without sedative or muscle relaxant effects such as Zolpidem, likewise the model was able to suggest changes that would not work.

This type of methodology will allow suggesting the structural changes that optimize the desired pharmacological activity.

1. Vlainić, J, Pericić, D. (2010) *Brain Res.* 15(1310), 181-188.

ANDSYSTEM: AUTOMATED RECONSTRUCTION OF GENE NETWORKS FOR OMICS-DATA INTERPRETATION IN MEDICAL AND BIOLOGICAL RESEARCH

***V.A. Ivanisenko*[1,2,3], *P.S. Demenkov*[1,2], *T.V. Ivanisenko*[1,2], *N.V. Ivanisenko*[1,2], *E.A. Antropova*[1], *A.V. Adamovskaya*[1,3], *P.M. Revva*[1,3], *I.N. Lavrik*[1,4], *N.A. Kolchanov*[1,3]**

[1]*Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*
[2]*Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*
[3]*Novosibirsk State University, Novosibirsk, Russia*
[4]*Translational Inflammation Research, Medical Faculty, Center of Dynamic Systems, Otto von Guericke University Magdeburg, Magdeburg, Germany*

Currently, the most popular approach to analyze omics data is to search for overrepresented GO biological processes or KEGG metabolic pathways. However, such an analysis does not provide information on molecular genetic interactions between genes that describe the mechanisms of observed perturbations in transcriptomic, proteomic, and metabolomic profiles. Gene networks have found wide application for the reconstruction of molecular genetic mechanisms. Gene networks are a graph of knowledge about molecular genetic interactions of various nature occurring in living systems. The main source of data for the reconstruction of gene networks are scientific publications. The number of scientific publications in the field of biomedicine is more than 1 million every year. At the same time, the number of articles published per year is steadily growing. Therefore, to analyze such a large amount of scientific literature, methods of automatic analysis of texts in natural language (NLP) are used.

We have developed a cognitive software and information system ANDSystem, designed to automatically extract knowledge from the texts of scientific publications and reconstruct gene networks on this basis [1, 2, 3]. ANDSystem has been used to solve a wide range of problems, including interpretation of omics data (genomic, transcriptomic, proteomic, metabolomic, and epigenomic), identification of disease biomarkers, search for pharmacological targets, drug repurposing, etc.

Reconstruction of signaling pathways for the regulation of cellular biological processes by viral proteins can help in the search for new pharmacological targets. Using the ANDSystem, the associative gene networks describing the potential regulation of the apoptosis process by HCV viral proteins were reconstructed. Another direction of research was related to the study of the molecular mechanisms of pathological processes in hepatocellular carcinoma. Induction of apoptosis by the cell response to mechanical stress caused by tumor tissue compaction is important antitumor mechanism. The analysis of gene networks made it possible to identify potential signaling pathways linking genes involved in the cell response to mechanical stress with key genes of the external pathway of apoptosis. Potential molecular mechanisms of dysfunction of these signaling pathways in hepatocellular carcinoma have also been reconstructed.

The work on the ANDSystem development was financially supported by the budget project No. FWNR-2022-0020. The reconstruction and analysis of gene networks was supported by the ERA-NET project "Target identification and drug development in liver cancer (TAIGA) " (agreement with the Ministry of Education and Science of Russia No. 075-15-2021-944).

1. Ivanisenko, T.V., Saik, O.V., Demenkov, P.S., Ivanisenko N.V., Savostianov A.N., Ivanisenko V.A. (2020) *BMC Bioinformatics*, 21, 228.
2. Ivanisenko, V.A., Demenkov, P.S., Ivanisenko, T.V., Mishchenko, E.L., Saik, O.V. (2019) *BMC Bioinformatics*, 20(Suppl. 1), 34.
3. Ivanisenko, V.A., Saik, O.V., Ivanisenko, N.V., Tiys, E.S., Ivanisenko, T.V., Demenkov, P.S., Kolchanov, N.A. (2015) *BMC Systems Biology,*. 9(Suppl. 2), S2.

COMPUTATIONAL CHARACTERIZATION OF N-ACETYLASPARTYLGLUTAMATE SYNTHETASE: FROM THE PROTEIN PRIMARY SEQUENCE TO PLAUSIBLE CATALYTIC MECHANISM

***A. Kniga*[1,2], *I. Polyakov*[1,2], *A. Nemukhin*[1,2]**

[1]*Department of Chemistry, Lomonosov Moscow State University, Moscow, Russia*
[2]*Emanuel Institute of Biochemical Physics RAS, Moscow, Russia*

The aim of the present study was to apply methods of bioinformatics to obtain the structure of N-acetylaspartylglutamate synthetase (NAAGS) and to characterize dynamics of its enzyme-substrate (ES) complex using the supercomputer molecular modeling. The structure of this protein was not yet experimentally determined, although it realizes the final second step in the metabolic pathway leading to the most abundant neuropeptide in the human brain.

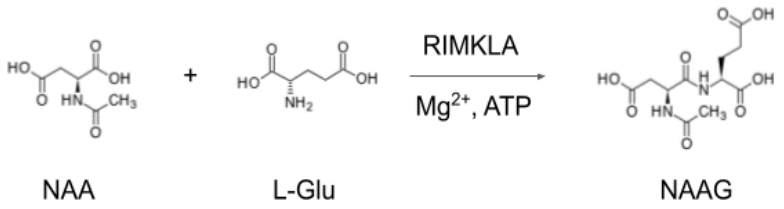


Figure 1. Condensation of N-acetylaspartate and L-Glu is catalyzed by ligase NAAGS, product of RIMKLA gene.

Multiple sequence alignment was constructed by searching sequence of a human gene RIMKLA, coding NAAGS [1], against protein sequence databases using MMseqs2. The structure of NAAGS was obtained using the pre-trained AlphaFold2 [2] model. C-terminal disordered region (305-391) was trimmed. The reactants (ATP, N-acetylaspartate and L-glutamate) were deposited into the active site cavity using the structural superposition onto the homologous protein structures from RCSB PDB.

Subsequent structural refinement of the constructed ES complexes was carried out with the help of large-scale classical molecular dynamics calculations. On the top of the construct, molecular dynamics simulations with the quantum mechanics/molecular mechanics interaction potentials were performed with NAMD/TeraChem [3] for the most promising ES conformations. Analysis of the latter allows us to propose a plausible catalytic mechanisms of chemical reactions in the enzyme active site. The applied computational strategy opens the way towards *ab initio* enzymology using modern supercomputer simulations.

This study was supported by the RSF project No. 18-13-00030. The calculations were carried out using equipment of the shared research facilities of computing resources at Lomonosov Moscow State University.

1.Collard, F. et al. (201) *J. Biol. Chem.*, 285(39), 29826-29833.
2. Jumper, J. et al. (2021) *Nature*, 596, 583-589.
3. Melo, M., Bernardi, R., Rudack, T. et al. (2018) *Nat. Methods*, 15**,** 351-354 (2018).

DE NOVO GENERATION OF SYNTHETICALLY FEASIBLE MOLECULES

G. Minibaeva*, *A. Ivanová*, *P. Polishchuk

Institute of Molecular and Translational Medicine, Faculty of Medicine and Dentistry, Palacký University, Czech Republic

Finding molecules that selectively interact with a biological target is a key step in drug discovery. Computer-aided drug design methods are nowadays widely used in the process of drug discovery and development. *De novo* molecular design approach for the search for new biologically active molecules is increasingly used. The application of this approach leads to new compounds with desired pharmacological properties that constitute novel intellectual property. One of the major limitations of *de novo* drug design tools is the synthetic accessibility of the generated molecular structures.

In the current work, a tool for *de novo* growing of drug-like compounds within a binding site of a target protein was developed. This tool includes the creation of molecules by concatenating fragments with the dynamic evaluation of a new molecule. To generate ligand structures we use the chemically reasonable mutations approach (CReM) [1] and to assess their binding to a target protein we use molecular docking by AutoDock Vina [2]. For generating molecules by CReM we use a database of interchangeable fragments, which was prepared from a data set of known compounds. Following the main idea of CReM, generation of new structures (replacement of fragments with other ones) is possible if fragments are in the same local chemical environment of a given radius. This combination (fragments of existing molecules + local chemical context) makes generated structures more synthetically feasible than simple linking of fragments like it is usually performed.

Two modes of the developed tool are available to create molecules. In the former case, this is iterative growing of a fragment co-crystallized with a protein preserving the position of the parent part of the molecule (hit expansion). In the latter case we use a preliminary created set of starting fragments from ChEMBL compounds to initialize *de novo* generation. Those fragments have from 8 to 15 heavy atoms, from 1 to 5 distinct hydrogen-bond donor/acceptors centers, logP < 2, TPSA > 25A^2, at most one halogen atom, at most two rotatable bonds and no structural alerts. This starting set of fragments is docked and iteratively grown. We implemented several strategies to select molecules on each iteration: greedy, Pareto and clustering-based selection.

To illustrate both hit expansion and *de novo* design the developed tool was applied to grow small ligands co-crystallized with 3C-like protease of SARS-CoV-2 (5RGX, 5RH2) and to *de novo* generation of ligands of CDK2 and other targets frequently used in benchmarking studies. During testing of the tool, it was studied how the choice of the following parameters such as fragment databases and the context radius affected the structural diversity and synthetic accessibility of generated compounds. Based on obtained results, we concluded, the synthetic complexity of the generated structures decreases with increasing radius, as well as with using a base of fragments obtained from more synthetically accessible compounds.

This work was funded by the INTER-EXCELLENCE LTARF18013 project (MEYS), the European Regional Development Fund - Project ENOCH (No. CZ.02.1.01/0.0/0.0/16_019/0000868) and ELIXIR CZ research infrastructure project (MEYS Grant No: LM2018131).

1. Polishchuk, P. (2020) *Journal of Cheminformatics*, 12(1), 1-18.
2. Trott, O., Olson, A.J. (2010) *Journal of Computational Chemistry*, 31(2), 455-461.

A COMPREHENSIVE COMPUTATIONAL PHARMACOKINETICS IDENTIFICATION OF BIOTRANSFORMED LEADS FROM *CURCUMA CAESIA* ROXB

***K.S. Mukunthan*[1], *S. Balaji* [1], *N.P. Trupti*[2]**

[1]*Department of Biotechnology, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, India*

[2]*Department of Integrative Biology, Vellore Institute of Technology, Vellore, India*

The biotransformation of drug happens with various metabolic stages and most of the investigations are focused on metabolizing enzymes such as CYP because it is the entry point of interaction for any drug candidate that enters the liver. Most predictions of ADMET, consider the inhibition of Cytochrome P450 with lead structures at the atomic level with the models of human Cytochrome P450. To date, predictions of monooxygenase based chemical transformation of the drug candidates with the metabolizing enzymes are still in primitive levels of research. In the current study, *in silico* screening [103 (native compounds) + 103 (Reduced form) + 103 (oxidized form)] of 309 exclusive compounds of *C. caesia* was carried to identify lead molecule. *In silico* absorption, distribution, novel metabolic prediction and toxicity resulted in the discovery of lead molecule i.e. Compound No 3. Further molecular docking assessment of the metabolically biotransformed leads identified using simplified pharmacophore analysis showed good inhibition potential against Peptidyl-prolyl cis-trans isomerase (PIN) and thus promises to be a good cancer inhibitor. The cost-effective approach presented in this paper could be used to filter toxic compounds from the drug discovery cycle.

1. Zou, Y., Li, H., Graham, E.T. et al. (2020) *Nat. Chem. Biol.* 16, 302–309.
2. Lee, S.K., Lee, I.H., Kim, H.J., Chang, G.S., Chang, J.E. (2003). The PreADME Approach: web-based program for rapid prediction of physic-chemical, drug absorption and drug-like properties. Blackwel Publishing, Boston, MA. 418-420.

DEVELOPMENT OF BIOMEDICAL EDUCATIONAL PROGRAMS

***Y.L. Orlov*[1,2], *I.A. Volkov*[1], *S.Y. Simonenko*[1], *A.A. Anashkina*[1,3]**

[1]*I.M. Sechenov First Moscow State Medical University of the Ministry of Health of the Russian Federation (Sechenov University), Moscow, Russia; y.orlov@sechenov.ru;*
[2]*Agrarian and Technological Institute, Peoples' Friendship University of Russia, Moscow, Russia;*
[3]*Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia*

Development of computer programs for drug discovery needs qualified scientific staff, not only software development and technical support. Growth of biomedical data volume and availability of new experimental technologies give solid background for complex computer modeling limiting rather by human resources than hardware. Bioinformatics education faces new challenges related to changing educational standards, distant education on online meeting formats.

Scientific challenges in connection with the coronavirus pandemic have raised a number of research and educational problems, changes in the methodology for mastering scientific disciplines. Medical universities use new E-health technologies, one of the global areas of which is telemedicine. Medical teleconsultations make it possible to increase the availability of medical care for the population of remote areas, elderly and inactive patients, which is especially relevant for monitoring the spread of coronavirus infection. The First Moscow State Medical University of the Ministry of Health of Russia (Sechenov University) and the Institute of Digital Medicine deal with the problems of digitalization of medicine, provide a platform for discussing existing issues in the development of medical technologies, online conferences, and new educational programs. The Russian Journal of Telemedicine and E-Health continues series of publication on this topic (https://jtelemed.ru/). We have arranges series of international journal issues on gene expression regulation as well. Participation of the students in the journal publications serves as good practice. It was realized in recent joint publications on gene network models and gene expression regulation co-authored by the students [1, 2]. Distinct features for the research is methodical part relying only on the bioinformatics web-server and online-tools such as PANTHER, DAVID, GeneCards, STRING-DB, GenMANIA ([2]).

Gene expression regulation at the transcriptome, genome, cell, and tissue levels is a complex phenomenon demanding the development of bioinformatics tools. Molecular mechanisms of human disease progression as well as gene expression in laboratory animal models are being studied using sequencing approaches coupled with advanced analytics. Modern computational approaches depend on the ability to reconstruct gene networks and model the protein structure, as it was reviewed recently. The analytic techniques were discussed at the BGRS (Bioinformatics of Genome Regulation and Structure/Systems Biology) biannual computational biology meeting in Novosibirsk, highlighting recent advances in the evolution, biomedicine, and biotechnology areas. We have organized special journal issues in the field of bioinformatics of gene expression in Frontiers in Genetics, PeerJ (https://peerj.com/collections/72-bgrs-sb-2020), and then in International Journal of Molecular Sciences [3]. The collections of papers show insights into the fields of genomics, transcriptomics, and proteomics, as well as some works conducted in model organisms.

This study was supported in part by RSF grant (A.A.). The publication was prepared with the support of the RUDN University Strategic Academic Leadership Program (recipient Y.O.).

1. Orlov, Y.L., Galieva, A.G., Orlova, N.G. et al. (2021) *Biomeditsinskaya Khimiya*, 67(3), 222-230.
2. Gubanova, N.V., Orlova, N.G., Dergilev, A.I. et al. (2021) *Journal of Integrative Bioinformatics*, 18, 20210031.
3. Anashkina, A.A., Leberfarb, E.Y., Orlov, Y.L. (2021) *Int. J. Mol. Sci.* 22, 12146.

GENE NETWORK ANALYSIS FOR COMPLEX DISEASE USING ONLINE BIOINFORMATICS TOOLS

R.R. Shakirova[1], N.G. Orlova[2,3], M.G. Elistratova[1], A.V. Korobeynikova[1], Y.A. Mozyleva[1], Y.L. Orlov[1,4]

[1]I.M. Sechenov First Moscow State Medical University of the Ministry of Health of the Russian Federation (Sechenov University), Moscow, Russia
[2]Financial University under the Government of the Russian Federation, Moscow. Russia
[3]Moscow State Technical University of Civil Aviation, Moscow, Russia
[4]Agrarian and Technological Institute, Peoples' Friendship University of Russia, Moscow, Russia

Development of computer pipelines for drug discovery could be based on online bioinformatics tools. Functional annotation of genes and gene ontology analysis for complex diseases using available bioinformatics tools provide background to search new targets for therapy. We aimed develop computer pipelines to study set of complex heterogeneous disease such as cancers and mental disorders using online bioinformatics tools. Examples of gene lists studied are related to glioma, glioblastoma, cystic fibrosis, Kaposi's sarcoma, Parkinson's disease and dementia. Thus, cystic fibrosis is a systemic hereditary disease caused by a mutation in the gene for the transmembrane regulator of cystic fibrosis. Parkinson's disease and dementia present socially important complex diseases. Functional annotation of genes related to the disease could be retrieved based on genetic databases and cross-validated by integrating complementary experimental data. Gene network reconstruction for a set of genes (proteins) proved to be effective approach to study mechanisms underlying disease progression [1-4].

We used online bioinformatics tools such as GeneCards (https://genecards.org), GeneMANIA (https://genemania.org/) and STRING-DB (https://string-db.org/) for annotation of gene list for the disease, reconstruction of gene network and comparative analysis of gene ontology categories. The Internet-resource OMIM (Online Mendelian Inheritance in Man) (https://omim.org/) and GeneCards suit (GeneCards.org) were used for gene list reconstruction. DAVID (https://david.ncifcrf.gov/) and PANTHER resources (Protein Analysis THrough Evolutionary Relationships) (http://pantherdb.org/) were used to analyze the categories of gene ontologies. We reconstructed gene networks for series of oncological diseases - glioma, lymphoma, and Kaposi's sarcoma. The work on gene network models include glioma [2], Parkinson's disease [1], and metabolic syndrome [4]. The steps include collection of a list of genes associated with the development of the disease (e.g., Parkinson's disease), analyze gene ontology categories for such a list, and reconstruct the gene network. For the key disease genes derived from the gene network structure analysis, drug search options are considered. This approach has already been tested for teaching the course "Fundamentals of Bioinformatics and Databases Management" at the Sechenov University (Gubanova et al., 2021; Dergilev et al., 2021).

The study is supported by the Potanin Foundation grant for masters' teachers (ГК22-000797).

1. Orlov, Y.L., Galieva, A.G., Orlova, N.G. et al. (2021) *Biomeditsinskaya Khimiya*, 67(3), 222-230.
2. Gubanova, N.V., Orlova, N.G., Dergilev, A.I. et al. (2021) *Journal of Integrative Bioinformatics*, 18, 20210031.
3. Dergilev, A.I., Orlova, N.G., Dobrovolskaya, O.B., Orlov, Y.L. (2021) *Journal of Integrative Bioinformatics*, 18, 20200036.
4. Tiis, R.P., Osipova L.P., Galieva, E.R. et al. (2021) *Biomeditsinskaya Khimiya*, 67(3), 213-221.

TCSTF: TOOL FOR CATEGORIZATION OF SHORT TEXT FRAGMENTS

P.V. Pogodin**, E.V. Ilgisonis, O.A. Tarasova, O.I. Kiseleva, D.A. Filimonov, E.A. Ponomarenko.

Institute of Biomedical Chemistry, Moscow, Russia

Tabular data covering biomedical subjects are available for the researchers in the field of artificial intelligence and machine learning in particular to be used for modeling. Nevertheless, raw texts still cannot be ignored: biological assays' descriptions, patients' comments on the usage of therapeutics, respondents' free answers – contain useful information, which is complementary to the records and labels given in the tabular form. In this article we describe the Tool for Categorization of the Short Text Fragments (TCSTF). TCSTF was developed to simplify the fascinating, but laborious text analysis and categorization (labeling) procedure, which in turn will help the researchers to model, predict and explain biomedical phenomena better. Examples provided in this study: rectification of the preliminary automated vocabulary-based categorization of text fragments describing studies of proteins' function (Open Targets and neXtProt data) and descriptions of assays, in which antifungal activity for the pairs of chemicals had been presumably measured, demonstrated the applicability of TCSTF to this task and its relevancy to the current state of affairs in biomedical and (Q)SAR modeling using databases. Describing the TCSTF we expect to interest and engage the researchers both from the fields of artificial intelligence and biomedicine to use TCSTF or similar tools, while transforming the available data in their completeness into the valuable biomedical knowledge. We develop TCSTF using R language, its libraries for the scientific computing and web technologies.

This study was performed in the framework of the State Academies of Sciences Basic Research program for 2021-2030 (No. 122030100170-5).

SARS-COV-2 MAIN PROTEASE INHIBITION WITH CARMOFUR: A COMPUTATIONAL STUDY

***I.V. Polyakov*[1,2], *B.L. Grigorenko*[1,2], *A.V. Nemukhin*[1,2]**

[1]*Department of Chemistry, Lomonosov Moscow State University, Moscow, Russia*
[2]*Emanuel Institute of Biochemical Physics RAS, Moscow, Russia*

Proposal of efficient inhibitors of the SARS-CoV-2 enzymes is an important area of research due to the continuous COVID-19 pandemic [1]. We report the results of modeling the interaction of the SARS-CoV-2 main protease (Mpro) with its covalent inhibitor carmofur (Fig. 1) using the combined quantum mechanics and molecular mechanics (QM/MM) approach. Calculations were carried out to determine the reaction energy profile and structures of the reactants, the product, and the transition state/intermediate. Two software packages, NWChem [2] and Q-Chem [3] were applied to perform QM/MM simulations.

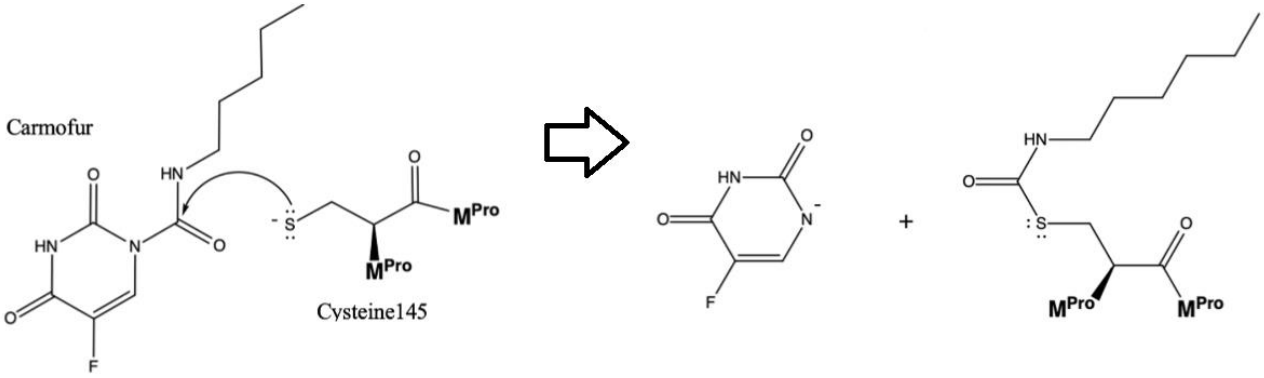


Figure 1. The reaction between carmofur and Mpro considered in QM/MM calculations.

We found that details of QM/MM simulations, such as the treatment of charges in the QM/MM boundary region or details of the QM/MM electrostatic embedding scheme should be fully characterized in order to ensure the reproducibility of the results. Our study reveal that the use of different QM/MM computational protocols implemented in different software packages may lead to certain discrepancies, because many parameters are hard-coded and cannot be changed via input files.

On the positive note, these benchmark calculations lead to the consistent results, namely, to the close energy profiles and structures of the located stationary points. Thus, we conclude that prediction of prospective covalent inhibitors for the Mpro and other enzymes can be successfully accomplished by properly documented QM/MM modeling.

This study was supported by the Russian Science Foundation (project No. 19-73-20032). The authors acknowledge collaboration with Prof. A.I. Krylov, Prof. S. Faraji and G. Giudetti in this research. The calculations were carried out using the equipment of the shared research facilities of computing resources at Lomonosov Moscow State University and of the resources of the Joint Supercomputer Center of the Russian Academy of Sciences.

1. Narayanan, A. et al. (2022) *Commun. Biol*, 5, 169.
2. Aprà, E. et al. (2020) *J. Chem. Phy.s*, 152, 184102.
3. Shao, Y. et al. (2015) *Mol. Phys.*, 113, 184–215.

SPATIOTEMPORAL IDENTIFICATION OF BINDING SITES WITH COMPUTER VISION

***P. Popov*[1], *I. Kozlovskii*[1]**

[1]*iMolecule, Skolkovo Institute of Science and Technology, Moscow, Russia*

Identification of novel macromolecular binding sites expands druggable genome and opens new opportunities for drug discovery. Generally, a binding site is a structural property of a macromolecule, that can be present or absent depending on the three-dimensional conformation of the molecule. Furthermore, local binding site properties determines type of a ligand binding to it. For example, small molecule and peptide binding sites differ in geometrical and physicochemical properties; these emphasise importance of careful dataset and benchmark collection in order to assess performance of the binding site prediction tools. On the other hand, binding site identification resemble the object detection problem in 3D computer vision. Indeed, one can consider macromolecular conformations as 3D-images, binding sites as objects on these 3D images to detect, and conformational ensembles of macromolecules as 3D-videos to analyze. In such a setup different types of binding sites correspond to different classes of objects to detect. Using 3D convolutional neural networks we developed BiteNet's predictive models for the binding site identifications [1, 2, 3]. More precisely, we firstly collected binding site type-specific datasets for protein-small molecule, protein-peptide, and nucleic-small molecule binding sites. Then, we applied sequence- and structure-based clusterization for rigorous train-validation-test splits. Finally, we trained the models using as the loss function the distance between the truth and predicted geometrical center of a binding site. We observed superior average precision with respect to the state-of-the-art approaches. In a retrospective case study, we considered the oncogenic G12C variant of K-Ras - one of the first oncogenes discovered, which remained elusive for druggability for decades. For a long time the orthosteric nucleotide binding site was considered as the main target in the structure-based drug discovery campaigns. However, allosteric binding sites in K-Ras, named Switch II, and Switch I/II pockets, were recently observed [4, 5]. Due to the highly flexible nature of the K-Ras protein structures, such binding sites are likely to be overlooked in the *apo* 3D conformations. We used full-atomic long-range molecular dynamics simulations of K-Ras in water to obtain representative conformational ensemble of K-Ras. We then analyzed the obtained simulation trajectories using the derived predictive models for the binding site identification and observed the nucleotide, Switch II, and Switch I/II pocket, as the top-scored predictions. Notably, for a fair retrospective study we excluded K-Ras and its homologous from training. Interestingly, we observed that while the nucleotide binding site is detected along the entire simulation, the allosteric binding sites were identified in ~25% of simulations. This can be explained by the transient nature of the binding sites, for which ligands are required to stabilize the bound conformations.

This study was supported by the Russian Science Foundation (RSF) research grant No. 22-13-00435.

1. Kozlovskii, I., Popov, P. (2020) *Communications biology*, 3(1), 1-12.
2. Kozlovskii, I., Popov, P. (2021) *Journal of chemical information and modeling*, 61(8), 3814-3823.
3. Kozlovskii, I., Popov, P. (2021). *NAR genomics and bioinformatics*, 3(4), lqab111.
4. Maurer, T. et al. (2012) *Proceedings of the National Academy of Sciences*, 109(14), 5299-5304.
5. Ostrem, J. M. et al. (2013) *Nature*, 503(7477), 548-551.

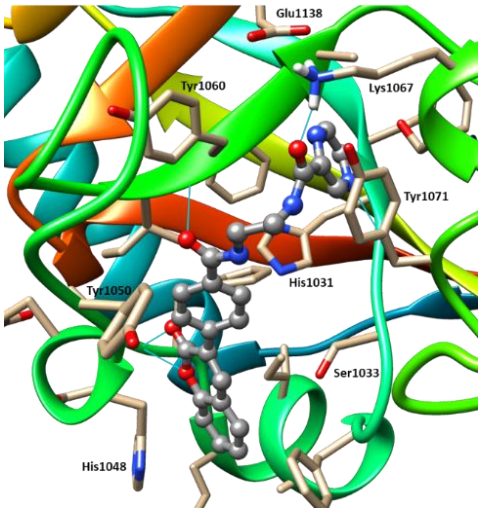
DISCOVERY OF NOVEL TANKYRASE INHIBITOR CHEMOTYPES: AN INSIGHTFUL TEST CASE FOR VIRTUAL SCREENING AND MOLECULAR MODELING APPROACHES

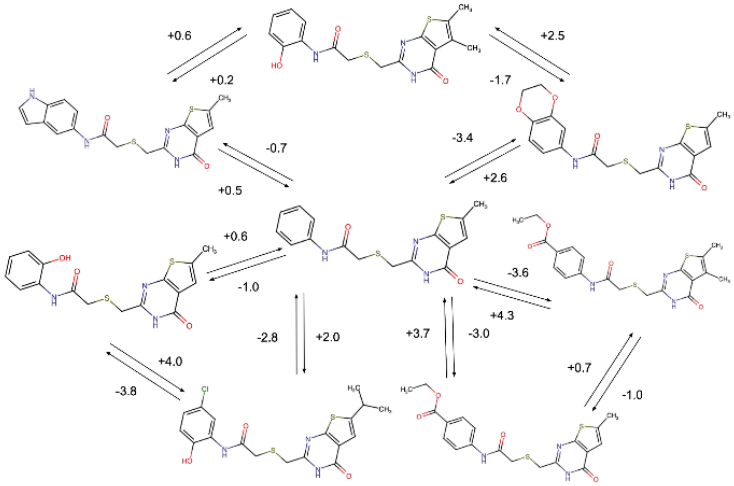
***E.V. Radchenko*[1], *V.P. Berishvili*[1], *A.N. Kuimov*[2], *A.E. Voronkov*[1,3], *V.A. Palyulin*[1]**

[1]*Department of Chemistry, Lomonosov Moscow State University, Moscow, Russia*
[2]*A.N. Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, 119991, Moscow, Russia*
[3]*Digital BioPharm Ltd., Oslo, Norway*

Tankyrase enzymes (TNKS), members of the poly(ADP-ribose) polymerase (PARP) superfamily, control the Wnt pathway and represent a promising target in the search for potential anti-cancer agents that can also be used in synergic combinations with drugs targeting other pathways. Using the molecular docking and machine learning-based virtual screening techniques along with the physico-chemical and ADMET profile prediction as well as molecular dynamics simulations, we have identified a number of candidate compounds in a subset of the ZINC database. Out of 36 compounds biologically evaluated *in vitro* for their inhibition of the TNKS2 enzyme using immunochemical assay, 13 compounds belonging to three different chemotypes have shown various levels of inhibitory activity. For two compounds, the IC_{50} values below 10 nM and 30 nM were obtained.

In retrospective analysis of the results, it was found that, although quite useful in virtual screening, the relatively simple scores based on molecular docking (even with target-specific machine learning-based scoring functions) or MM-PBSA methods proved unsuitable [1] for predicting the effect of structural modification or for accurate ranking of the compounds based on their binding energies. On the other hand, the molecular dynamics simulations and Free Energy Perturbation (FEP) calculations allowed us to further decipher the structure-activity relationships. In matching molecular pairs or networks of congeneric compounds, the Relative Free Energy Perturbation (RFEP) technique enables efficient activity ranking. These approaches can be applied at the subsequent lead optimization stages.





This study was supported by the Russian Foundation for Basic Research (project No. 18-515-80028) under the BRICS STI cooperation program.

1. Berishvili, V.P., Kuimov, A.N., Voronkov, A.E., Radchenko, E.V., Kumar, P., Choonara, Y.E., Pillay, V., Kamal, A., Palyulin, V.A. (2020) *Molecules*, 25, 3171.

THE PHILOSOPHY AND PROSPECTS OF FRAGMENT CONTRIBUTION ESTIMATIONS IN DRUG DISCOVERY

***D. Shulga*[1,2], *N. Ivanov*[1], *V. Palyulin*[1]**

[1]*Lomonosov Moscow State University, Chemistry Dept., Moscow, Russia*
[2]*molmodel.com, Moscow, Russia*

Contemporary drug discovery faces hard times caused by stronger regulatory requirements and exhausting of "the low hanging fruits" of blockbusters' targets. The systemic response of the pharmaceutical industry is the broadening of the Open Innovation style wide net of cooperation and vivid adoption of Agile approaches, that encourage fast and lean hypothesis testing during the drug design process. Definitely relevant in silico means which help to achieve the above goals are necessary.

Fragment in molecule is a very natural and still vague and disputable conception, which despite the lack of strict scientific definition has been widely used in everyday practice. However this use has been mostly restricted to qualitative applications. The focus of our study is to investigate the scope in which the decomposition of affinity in terms of separate contributions of different ligand fragments provides useful and lean means, which can boost drug discovery process.

Conceptually we derive our approach from the same principles that were laid into foundation of the well-known and appreciated fragment-based drug discovery (FBDD) approach. Moreover, we consider our approach as a reverse version of FBDD conception, so we called it reverse fragment-based drug discovery (R-FBDD) [1]. However we have found that the perception of the two approaches, despite stemming from the same root, differs significantly, thus forming a phenomenon, which we call *fragment-based perception hysteresis*. This phenomenon, as well as its both true and false pros and cons will be presented.

In the second part the recently proposed practical implementation of in silico fragment contribution estimation using scoring functions [1, 2] is employed to exemplify the potential practical use cases, where the proposed approach provides the researchers with additional valuable information and guidance. For instance, our method can be used to rationally decide on which substituents are superfluous in a moderate activity hit compound, this use case being very common in lean practice. In this regard the group efficiency (GE) metrics for a certain group, which is a direct analog of the ligand efficiency (LE) for the entire ligand, can be used to guide decision to get rid of effectively ballast substituents. This move opens the opportunity of introduction of more effective substituents and thus leading not only to more active/affine ligand, but also to a structure with better ligand efficiency (LE). Other useful examples and use cases will be also presented.

The proposed approach introduces additional dimensions, which can be used to rank different substituents of the core structure. Moreover the value of the central scaffold can be critically evaluated in order to make better judgments on how much further development using this scaffold is risky.

The performed analysis reveals that the in silico fragment contribution approach is a promising new tool, which has deep thinking tradition behind it. Thus it is natural for researchers both to analyze and to design in terms of fragments and their properties. An important property of our proposed approach is its ease of application, which renders it as a valuable tool for lean and agile modern drug discovery, which heavily and wisely relies on in silico methods to boost the entire drug discovery process.

This work has been supported by the Russian Science Foundation, grant No. 22-23-00729.

1. Shulga, D. A., Ivanov, N. N., and Palyulin, V. A. (2021) Mend. Comm. 31(3), 291–293.
2. Shulga, D. A., Ivanov, N. N., and Palyulin, V. A. (2022) Molecules 27(6), 1985.

SELF CONSISTENT CLASSIFIER SAR APPROACH

L. Stolbov, D. Filimonov, V. Poroikov

Institute of Biomedical Chemistry, Moscow, Russia

Ligand-based drug design is used in various drug discovery projects to facilitate the identification of possible hits and reduce the ever-increasing costs of finding new chemical entities. Based on this approach, the analysis and modeling of Structure-Activity Relationships (SAR) are widely used for *in silico* studies and virtual screening. The SAR approach significantly depends on the available experimental data utilized as the training set. To overcome the issues with data quality and diversity, classification methods are used to designate active and inactive structures. Though a number of algorithms have been developed to satisfy the activity classification needs, the ability of the model to produce a generalized predictions is still challenging.

In this study, a new method for classifying chemical compounds by their activity, based on the statistical regularization, is presented, which we previously applied for self-consistent regression (SCR) [1, 2]. Quantitative neighborhood of atoms (QNA) molecular descriptors, which have been implemented earlier in our "Star Track" approach and GUSAR software [3, 4], are used. The main feature of SCR is that based on a strict mathematical basis, and the automatic self-consistent selection of molecular descriptors with the best correspondence to the analyzed SAR is executed. Classification problem strictly corresponds to a Bernoulli trial, from which the logistic regression is derived. Based on SCR, we have developed a logistic SCR (LSCR) algorithm and its modification, called exponential SCR (ESCR). The peculiarity of ESCR is that in this variant of statistical regularization, the greatest importance is attached to the exclusion of classification errors. We illustrate the application of LSCR and ESCR to data on the anti-HIV compounds activity and toxicity data.

Data on anti-HIV activity were collected from different sources, including EBI ChEMBL [5] and NIAID HIV/OI/TB [6], toxicity data were taken from Tox21 [7]. The anti-HIV activity data were divided by chosen activity cutoff, the toxicity data were initially classified. The LSCR and ESCR algorithms are implemented from scratch using the *C++* programming language and integrated into the *R* environment with *Rcpp* mediation for results processing. The LSCR and ESCR classification results were compared with the SCR and the Support Vector Machine Classifier (SVM classifications for the same data. All models were validated using a 5-fold cross validation procedure.

The results of the work show the advantages of the LSCR and ESCR approaches in reducing the dimensionality of the constructed SARs while maintaining accuracy compared to the compared approaches. Thus, the difference between the characteristics of the models for the training set and those shown in validation studies decreases, which makes the subsequent predictions more fair.

This study was supported by the Russian Foundation for Basic Research grant No. 20-04-60285.

1. Filimonov, D.A., Akimov, D.V., Poroikov, V.V. (2004) *Pharmaceutical Chemistry Journal*, 38, 21.
2. Lagunin, A.A., Zakharov, A.V., Filimonov, D.A. et al. (2007) *SAR QSAR Environ. Res.*, 18, 285.
3. Filimonov, D.A., Zakharov, A.V., Lagunin, A.A. et al. (2009) *SAR QSAR Environ. Res*., 20, 679.
4. Zakharov, A.V., Peach, M.L., Sitzmann, M. et al. (2014) *J. Chem. Inf. Model.*, 54, 3, 713.
5. ChEMBL database URL [https://www.ebi.ac.uk/chembl/], Accessed on April 24, 2022.
6. NIAID HIV/OI/TB database URL [https://chemdb.niaid.nih.gov/], Accessed on April 24, 2022.
7. Tox21 public data URL [https://tripod.nih.gov//tox21/pubdata/], Accessed on April 24, 2022.

NEW INHIBITORS OF THE COAGULATION FACTOR XIIA: DOCKING AND EXPERIMENTAL VERIFICATION

A.V. Sulimov[1,2], D.C. Kutov[1,2], A.S. Tashchilova[1,2], I.S. Ilin[1,2], A.Yu. Potapov[3], D.Yu. Vandyshev[3], Kh.S. Shikhaliev[3], N.A. Podoplelova[4,5], M.A. Panteleev[4,5], V.B. Sulimov[1,2]

[1]Dimonta, Ltd, Moscow, Russia
[2] Research Computing Center of Lomonosov Moscow State University, Moscow, Russia
[3]Voronezh State University, Voronezh, Russia
[4]Russian Children's Clinical Hospital of the Pirogov Russian National Research Medical University, Moscow, Russia
[5]Center for Theoretical Problems of Physicochemical Pharmakology RAS, Moscow, Russia

The objective of the study is the search for low molecular weight inhibitors of the blood coagulation factor XIIa. This factor is among the most promising therapeutic targets for creating new generation of anticoagulants since deficiency of this factor protects against thrombosis without causing spontaneous bleeding [1]. The activation of the cascade of biochemical reactions is possible in two pathways: external (tissue factor pathway) and internal (contact) pathways. The position of coagulation factor XIIa in the cascade makes it possible to create a new class of anticoagulants that can block the internal pathway of coagulation activation without disturbing the chain of reactions of plasma hemostasis activated by the external pathway. Development of factor XIIa inhibitors is at an early stage, and the task of developing anticoagulants using such inhibitors is extremely important, especially taking into account an essential role of anticoagulants in the treatment of COVID-19. A virtual screening of candidates for inhibitors of factor XIIa is carried out using the SOL docking program [2] followed by the quantum-chemical calculations of the protein-ligand binding enthalpy using the PM7 method [3] and the COSMO solvent model implemented in the MOPAC program [4]. The database of organic compounds of the Voronezh State University, consisting of more than 19 thousand drug-like molecules, is used for screening. The best compounds selected by molecular modeling are tested *in vitro*, and several of them demonstrated inhibitory activity against factor XIIa at the micromolar level of IC_{50}. Some of these inhibitors demonstrate selectivity over the coagulation factors Xa and XIa. New inhibitors of factor XIIa found in the present work belong to various chemical classes and are perspective hits for the further development of the lead compounds for new generation of anticoagulants.

This research is financially supported by the Russian Science Foundation, Agreement No. 21-71-20031.

1. Kleinschnitz, C., Stoll, G., Bendszus, M., Schuh, K., Pauer, H.-U., Burfeind, P.; Renné, C., Gailani, D., Nieswandt, B., Renné, T. (2006) *J. Exp. Med.*, 203, 513–518.
2. Sulimov, V.B., Ilin, I.S., Kutov, D.C., Sulimov, A.V.(2020) *J. Turkish Chem. Soc. Sect. A Chem.*, 7, 259–276.
3. Stewart, J.J.P. (2013) *J. Mol. Model.*, 19, 1–32.
4. Stewart, J.J.P. Stewart Computational Chemistry. MOPAC2016 URL: http://openmopac.net/MOPAC2016.html (accessed on 8 April 2022).

SYNTHESIS, DOCKING, AND IN VITRO ANTICOAGULANT ACTIVITY ASSAY OF RHODANINE DERIVATIVES OF PYRROLO[3,2,1-*ij*]QUINOLIN-2(1*H*)-ONE AS NEW INHIBITORS OF FACTOR Xa AND FACTOR XIa

***A. Tashchilova*[1], *N. Novichikhina*[2], *A. Shestakov*[2], *I. Ilin*[1], *V. Sulimov*[1], *M.Panteleev*[3,4], *N. Podoplelova*[3]**

[1]*Voronezh State University, Voronezh, Russia*
[2] *Research Computing Center of Lomonosov Moscow State University, Moscow, Russia*
[3]*Russian Children's Clinical Hospital of the Pirogov Russian National Research Medical University, Moscow, Russia*
[4]*Center for Theoretical Problems of Physicochemical Pharmakology, Moscow, Russia*

Coagulation factor Xa and factor XIa are proven to be convenient and crucial protein targets for treatment for thrombotic disorders and thereby their inhibitors can serve as effective anticoagulant drugs. We have previously shown that effective inhibitors of these target proteins are possible in the series of derivatives of pyrrolo[3,2,1-*ij*]quinolin-2(1*H*)-one [1,2]. In the present work, we focused on the structure–activity relationships of rhodanine derivatives of pyrrolo[3,2,1-*ij*]quinolin-2(1*H*)-one and an evaluation of their activity against factor Xa and factor XIa. Design of inhibitors was carried out using docking followed by quantum chemical calculations of the enthalpy of protein-ligand binding. Docking was performed by the SOL docking program [3], and for the binding enthalpy calculations the PM7 semiempirical quantum-chemical method with the COSMO solvent model implemented in the MOPAC program were used. Docking-guided synthesis of more than 20 compounds based on tetrahydro-1*H*-pyrrolo[3,2,1-*ij*]quinolin-1-ylidene-2-thioxothiazolidin-4-ones was carried out. For the synthesis of new hybrid pyrrolo[3,2,1-*ij*]quinolin-2(1*H*)-one rhodanine derivatives, we used convenient structural modification by varying the substituents at the C8 positions and endo-N-atoms of rhodanine. In vitro testing revealed that four derivatives were able to inhibit both coagulation factors and three compounds were selective factor XIa inhibitors. An IC50 value of 3.68 μM for was found for the best factor Xa inhibitor and 2 μM for the best factor XIa inhibitor

The study was supported by the Russian Science Foundation grant No. 18-74-10097, https://rscf.ru/project/18-74-10097.

1. Novichikhina, N., Ilin, I., Tashchilova, A., Sulimov, A., Kutov, D., Ledenyova, I., Krysin, M., Shikhaliev, K., Gantseva, A., Gantseva, E. et al. (2020) *Molecules*, 25, 1889.
2. Medvedeva, S.M., Potapov, A.Y., Gribkova, I.V., Katkova, E.V., Sulimov, V.B., Shikhaliev, K.S. (2018) *Pharm. Chem. J.*, 51, 975–979.
3. Sulimov, V.B., Ilin, I.S., Kutov, D.C., Sulimov, A.V. (2020) *J. Turkish Chem. Soc. Sect. A Chem.*, *7*, 259–276.

MOLECULAR DOCKING-ASSISTED INVESTIGATION OF Cu(II) COMPLEXES CARRYING "SNS" PINCER-TYPE PYRIDINE-THIOETHER LIGANDS AS POTENTIAL DRUG CANDIDATES AGAINST SARS-COV-2

H.G. Sogukomerogullari[1], T. Taşkın Tok[2,3]

[1]Medical Services and TechniquesDepartment, Health Services Vocational School, Gaziantep University, Gaziantep, Turkey
[2]Department of Chemistry, Faculty of Arts and Sciences, Gaziantep University, Gaziantep, Turkey
[3]Institute of Health Sciences, Department of Bioinformatics and Computational Biology, Gaziantep University, Gaziantep, Turkey

The coronavirus disease 2019 (COVID-19) pandemic has posed a hazard to public health all over the world since the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) was discovered in late 2019. Since the start of the epidemic, scientists have put in a lot of effort in this field. However, the evaluation of the impact of newly synthesized drugs against coronavirus is lacking in these investigations. Examining newly synthesized compounds with a computer-assisted molecular docking study offers a significant benefit in terms of estimating and analyzing the biochemical activity and binding affinity of existing synthesized compounds against a biological target in a labor-, time-, and cost-effective manner.

Show, Van Koten, and Noltes characterized the structure of pincer type ligands and the pincer complexes they create in the 1970s as a result of their study. They noted that it is relevant to chemical research and a variety of other fields, as well as being effective in crucial investigations. [1,2] Pincer-type compounds have sparked a lot of attention due to their high activity in a variety of catalytic and biological activities [3,4]. Our aim in this study is to reveal with theoretical calculations whether the SNS-type pincer compound and its metal complexes will be effective against target proteins of SARS-CoV-2.

In this study, *in silico* investigations have defined and justified interaction processes between these molecules and Cu(II) at the atomic level. Furthermore, using molecular docking against COVID-19, the efficiency of the pincer ligands and their Cu (II) complexes produced was studied and discussed.

1. Moulton, C.J., Shaw, B.L. (1976) *J. Chem. SocDalton Trans.*, 11, 1020-1024.
2. Van Koten, G., Timmer, K., Noltes, J.G., Spek, A.L. (1978) *Chemical Communications*, 6, 249-250.
3. Sogukomerogullari, H. G., Yalçın, Ş. P., Ceylan, Ü., Aytar, E., Aygün, M., Richeson, D.S., Sönmez, M. (2019) *Journal of Chemical Sciences*, 131, 32.
4. Sogukomerogullari, H.G. (2022) *Computers in Biology and Medicine*, 145, 105512.

THE CONSENSUS ENSEMBLE MULTIDESCRIPTOR MULTITARGET NEURAL NETWORK MODELING OF PHARMACOLOGICAL ACTIVITY OF CHEMICAL COMPOUNDS

***P.M. Vassiliev*[1], *A.A. Spasov*[1], *M.A. Perfilev*[1], *D.V. Maltsev*[1], *D.A. Babkov*[1], *O.A. Luzina*[2], *A.N. Kochetkov*[1]**

[1]*Volgograd State Medical University, Volgograd, Russia*
[2]*Vorozhtsov Institute of Organic Chemistry, Novosibirsk, Russia*

The development of new approaches to the virtual screening of pharmacological compounds with a multiple mechanism of action or exhibiting nonspecific activity using artificial intelligence methods is one of the topical directions for increasing the efficiency of in silico search for new drugs.

The results of constructing classification consensus ensemble multidescriptor multitarget models of four different types of systemic pharmacological activity of chemical compounds (RAGE-inhibiting, anxiolytic, hypoglycemic and reducing LPS-intoxication) using the methodology of artificial perceptron feedforward neural networks are presented.

The training sets were formed according to the data on the structure and activity level of known experimentally studied substances available in international databases: 183 RAGE inhibitors; 216 anxiolytics; 318 hypoglycemic compounds; 12849 substances that reduce LPS intoxication. In all cases, three gradations of activity were considered: high, high or moderate, active.

The input neurons were docking energies in 34, 14, 20, and 7 biotargets, the most significant for the formation of the listed systemic effects, respectively. In the case of LPS inhibitors, 10 energies of boundary occupied MOs, 10 energies of boundary unoccupied MOs, and 92 types of QL descriptors of the 1st rank were added to 7 docking energies.

In all training sets, 7 sampling variables were added, which determine the options for the formation of training, test, and validation subsets.

A two-layer perceptron with a bottle-neck was chosen as the initial architecture of the neural network.

In total, ~ 100 000, ~ 125 000, ~ 135 000, ~ 2 500 neural networks were trained and four consensus ensemble multidescriptor multitarget neural network models were formed for the indicated types of activity. Each such model consists of three ensembles of neural networks for three gradations of activity (high, high or moderate, active), seven neural networks in each ensemble.

The average accuracy of the four models on three levels of activity for the simple consensus of the first level was 98.2 %, 98.8 %, 96.4 %, and 90.9 %, respectively.

The resulting models are used for directed search in silico for new systemic drugs with high activity: RAGE inhibitors for the treatment of complications in diabetes mellitus and Alzheimer's disease; anxiolytics of non-traditional action; multitarget hypoglycemic compounds for the treatment of type 2 diabetes mellitus; substances that prevent the development of hypercytokinemia in COVID-19.

The studies were supported by grants from Russian Foundation for Basic Research No. 18-015-00499, and No. 20-015-00164, Russian Science Foundation No. 20-13-00029, and Ministry of Education and Science of Russia No. 075-15-20-777.

FROM GPCR BASIC RESEARCH TO DRUG DISCOVERY VIA COMPUTATIONAL METHODS

***J. Liu*[1], *S. Wang*[2], *S. Yuan*[1,2]**

[1]*AlphaMol Science Ltd, Shenzhen, China;*
[2]*Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China*

G protein coupled receptors (GPCRs) are the most important drug targets for modern drug discovery. GPCRs are involved in different cell signalling including inflammation, neuron transmission, cell growth and many others. Almost all human diseases and disorders are related to GPCRs and their signalling. GPCRs have become popular targets for modern drug discovery and about 40% of the marketed drugs are targeting them. Drug discovery is a long and tedious process which costs at least 10 years and 2 billion USD. How to speed up this expensive process has become one of the most essential topics in pharmaceutical industry. With the progresses in both artificial intelligence and computational biology, advancing modern drug discovery via computational pharmacy plays more and more important roles.

The team of Dr. Yuan and his group has a long history in GPCR basic research as well as drug discovery. In 2014, the water channel activation mechanism was proposed by Dr. Yuan which has become one of the essential theories in the area of GPCRs. Following that, a systematic work was done for GPCRs including drug specificity, molecular switches, transmembrane movement features and many others. In December 2021, his group joint the global GPCR-DOCK 2021 contest. Three of the submitted models were ranked respectively as top 2, top3 and top 5 in the world for the kappa opioid receptor modelling and binding mode predictions. In this talk, Prof. Yuan will talk about how to apply basic molecular research to advance GPCR drug discovery. In addition, Dr. Yuan will also talk about his successfully story of how to advance "first-in-class" GPCR drug into clinical trials.

1. Zou, R., Wang, X. et. al. (2021) *Computational Molecular Science*, 12(2), E1565.
2. Bai, B., Zou, R., Chan, S., Li, H., Yuan ,S. (2021) *Molecules*, 26(15), 4625.
3. Hua, T., Li, X., et. Al (2020) *Cell*, 180(4), 655665.e18.
4. Liu, K., Wu, L., Yuan, S., et. al (2020) *Nature* 585, 135–140.
5. Chan, S., Wu, Y., Tan, L., Vogel, H. Cheng, J., Wu, D., Yuan, S. (2020) *ACS Central Science*, 6(2), 274–282.
6. Chan, S., Li, Y., Dahoun, T., Vogel, H., Yuan, S. (2019) *Trends in Biochemical Sciences*. 44(4), 312-330.
7. Chan, S., Wang, J., Palczewski, K., Filipek, S., Vogel, H., Liu, Z., Yuan, S. (2018) *Chemical Science*, 9, 6480-6489.
8. Yuan, S., Chan, H. C., Vogel, H., Filipek, S., Stevens, R.C., Palczewski, K. (2016) *Angewandte Chemie Int. Ed. Engl.*, 55, 35.
9. Yuan, S., Filipek, S., Palczewski, K., Vogel, H. (2014) *Nature Communications*, 5, 4733.

YOUNG SCIENTISTS CONTEST

BIOLOGICAL OR TECHNICAL NATURE OF TRANSCRIPTOMES HETEROGENEITY

V.A. Arzumanian*, *E.V. Poverennaya

Institute of Biomedical Chemistry, Moscow, Russia

Using the HeLa cell line as an example, it was shown that cells belonging to the same type can differ significantly in genomic and transcriptomic profiles [1]. At the same time, even in the case of good reproducibility in terms of gene expression, the molecular profile of transcript expression may differ significantly between samples. The purpose of this study is to evaluate the nature of transcriptome heterogeneity by comparing the various algorithms for calculating transcript expression using the widely cell lines as an example.

Four tools were selected for the study: Kallisto, Salmon (mapping-based mode and alignment-based mode) and eXpress. Each method was tested on 7 cell lines: HEK293, A549. K-562, HepG2, MCF7, Huh7 and Jurkat. RNA-Seq data were downloaded from the NCBI database matching to the following criteria: number of reads >20 million, average read length >50, performed by Illumina technology. Previously, cell lines were selected according to the reproducibility of gene expression.

As a result, tools with pseudo- and quasi-alignment methods showed a low correlation between transcript expression, in contrast to Salmon (alignment-based mode). Nevertheless, heterogeneity in the transcriptomic profile is observed for a part of samples, which requires a more careful study.

Thus, the transcripts abundance variance can be caused by both biological and technical differences in the case of using methods based on pseudo- and quasi-alignment.

This study was supported by RSF № 20-14-00328.

1. Y. Liu et al., (2019) *Nat. Biotechnol.*, 37, 314–322.

APPLICATION OF 2D-QSAR AND CHEMICAL READ-ACROSS ALGORITHM TO PREDICT THE ANDROGEN RECEPTOR BINDING AFFINITY

***A. Banerjee*, P. De, K. Roy**

Drug Theoretics and Cheminformatics Laboratory, Department of Pharmaceutical Technology, Jadavpur University, Kolkata, India

Endocrine Disruptor Compounds (EDCs) exist in the environment and promote adverse modifications of endogenous hormone regulation. By interfering with the body's endocrine system, these chemicals produce adverse developmental, reproductive, neurological, and immune effects in animals, abnormal growth patterns, and neurodevelopmental delays in children. Among these, certain compounds mimic the role of androgen which is responsible for controlling the development and maintenance of male sexual characteristics. In the present research, we have utilized the application of a two-dimensional quantitative structure-activity relationship (2D-QSAR) modeling technique to analyze the structural features of these chemicals responsible for binding to the androgen receptors (logRBA) in rats. We have collected the androgen receptor binding data from the EDKB database (https://www.fda.gov/science-research/endocrine-disruptor-knowledge-base/accessing-edkb-database). We have then employed the DTC-QSAR tool, available at https://dtclab.webs.com/software-tools for dataset division, feature selection, and model development. This tool is a complete package providing a user-friendly, easy-to-use GUI to develop regression or classification-based QSAR models involving variable selection techniques such as genetic algorithm and best subset selection. Dataset division was done by the Euclidean Distance approach method followed by feature selection using the Genetic Algorithm technique. The best descriptor combination selection for the pooled descriptors from the best GA derived models was done using the tool Best Subset Selection v2.1 available at https://dtclab.webs.com/software-tools. The final partial least squares (PLS) was evaluated using various stringent validation criteria. The developed model is robust, predictive, and should be a useful tool to predict the binding nature of EDCs to the androgen receptor. From the model, we interpreted that hydrophobicity in terms of octanol-water partition coefficient, aliphatic -CH group count, bulkiness of the structure, in addition to number of non-aromatic conjugated carbon atoms (sp2 hybridised), presence of CF3 group, percentage of Nitrogen present in the compounds contribute to the receptor binding affinity and thus increase toxicity, while presence of electron rich features like aromaticity in a molecule and presence of polar groups like alcohol, phenol or carboxyl groups decrease the receptor binding affinity and reduce toxicity. Additionally we have also performed chemical Read-Across using Read-Across-v2.0 available from https://sites.google.com/jadavpuruniversity.in/dtc-lab-software/home, and the results for the external validation metrics were found to be better in the Euclidean distance-based similarity considerations.

EXPLORING THE SCORING FUNCTION SPACE FOR STRUCTURE-BASED DRUG DESIGN

G. Bitencourt-Ferreira, W.F. de Azevedo, Jr.

School of Health and Life Sciences, Pontifical Catholic University of Rio Grande do Sul (PUCRS), Porto Alegre, Brazil

Molecular docking is one of the most employed computational techniques for structure-based drug design when the protein structure is available. This technique has two stages: determine the binding mode of the ligand on the active site of the protein and predict the binding affinity between these ligands, and protein. Both these stages are controlled by the scoring function in the docking program algorithm. Although most docking tools have robust and accurate algorithms to identify the binding mode of the ligand, the correct calculation of the binding affinity is still an open problem in the field. One of the reasons for this problem is the fact that all current docking programs make use of a classical scoring function, a predetermined theory-inspired equation that does not consider the specific attributes of our protein of interest. Having this problem in mind, we introduced the concept of the scoring function space. If we select a protein from the protein space [1], then it is possible to identify a set of ligands in the chemical space [2] that can interact with this protein. In that way, those two different spaces interact with each other. The scoring function space would be a third space, a mathematical space that comprises all the scoring functions. By exploring this space, it is possible to find a specific scoring function that can explain each interaction between the protein and chemical space. The idea of a specific scoring function considers the physical attributes in common from all the ligand interactions available for our protein of interest.

SAnDReS 2.0 is an open-source suite of programs developed by our lab that explores the scoring function space using machine learning methods. Our program creates protein-specific scoring functions with sixty-four different regression methods. The goal of this work is to use the SAnDReS 2.0 to develop a machine learning-based scoring function using docking results from the CDK2 protein and its inhibitors. We utilized the program Molegro Virtual Docker (MVD) to insert 77 inhibitors in the binding site of the CDK2. During molecular docking, the program calculates several different energy terms necessary for the two classical scoring functions that it is used. We employed SAnDReS 2.0 to import the molecular docking results and the energy terms calculated. Our specific scoring function is developed with six of the best energy terms calculated during the docking, and the weight of each term in the equation is determined with the regression methods of the program. The specificity of our scoring function is achieved with the division of our dataset, wherein we used a percentage of this dataset to train the equation and the rest to test the final scoring function.

The best scoring function was developed using the method *ExtraTreesRegressorCV*, the spearman correlation of the binding affinity was calculated by this scoring function, and the experimental binding affinity was 0.656 (p=0.0009). The prediction of the binding affinity improved significantly when we compared it with the results from the classical scoring functions of the MVD program. The Plants Score had a correlation of -0.298 (p=0.179) and the MolDock a correlation of -0.051 (p=0.820). With these results we propose that a protein-specific machine learning-based scoring function can guarantee a better prediction of the binding affinity than classical scoring functions. Our next step in this project is to compare our protein-specific approach to general machine learning-based scoring functions.

This study was financed by the CAPES (Finance Code 001), PRO-Stricto Scholarship Program – PUCRS, and CNPq (309029/2018-0).

1. Hou, J., Jun, S., Zhang, C. et al. (2005) *Proc. Natl. Acad. Sci. USA*, 102, 3651-6.
2. Lipinski, C., Hopkins, A. (2004) *Nature*, 432, 855-61.

INFORMATION EXTRACTION FROM TEXTS: ANTIVIRAL AGENTS ACTIVE AGAINST VIRUS OR HOST PROTEINS

N.Y. Biziukova, O.A. Tarasova, D.A. Filimonov, V.V. Poroikov

Institute of Biomedical Chemistry, Moscow, Russia

Due to active spread of viruses over the world, development of novel approaches to cure viral infections is essential. Understanding the mechanisms of interaction between human proteins and viruses provides an opportunity to find key host protein targets that are able to mediate dissemination of viruses in human body either directly or indirectly. Regulation of these interactions may allow, for example, reactivation of the virus in reservoirs and more successful elimination from the organism [1]. Although many such interactions are already identified, only small part of them used to develop drugs due to potentially diffuse toxic effects. In this regard, selection of related targets with less probability of toxic effects, as well as human proteins involved in the pathogenesis of viral diseases, is very important.

Growth of investigations dedicated to viral diseases and attempts to find new treatments lead to an extensive increase of scientific publications. Currently, comprehensive analysis of all available papers could not be performed manually. The main goal of our study is to develop an algorithm for extracting information from the text of scientific publications and test it in the case study for potential inhibitors of viral or human proteins involved in virus-host interaction for two viruses affecting immune system, human immunodeficiency virus (HIV) and hepatitis C virus (HCV).

We have collected automatically a set of publication abstracts relevant to (a) the mechanisms of virus-host interaction and (b) the inhibitory potency of chemical compounds on proteins involved in virus-host interactions using Python 3.10 scripts and requests to NCBI PubMed database. Extraction of the chemical and protein/gene named entities were performed using a combination of machine learning approaches based on conditional random fields (CRF) algorithm and naïve-Bayes approach, freely available as the web-services [2, 3]. Belonging of a protein or a gene to viruses or humans was also recognized using automatic queries in UniProt implemented in Python 3.10 scripts. For antiviral agents we collected additional information from ChEMBL 30. For relations' extraction we used rule-based approach. To implement the rule-based approach, we took a sample of 100 randomly selected publication abstracts and manually extracted pattern phrases which indicated any relation (i.e., "is the inhibitor of"). Analyzing texts, we found that part of pattern phrases usually connects specific named entities. To this end, this phenomenon was also turned into rules of order. All texts were splitted into sentences. For further relations' analysis, we considered only sentences that contain at least two recognized named entities. If two named entities were connected by a pattern phrase and satisfied discovered rules of order, we considered them as interacting.

Extracted relations were then verified manually considering the experimental data. Among the list obtained, for example, cyclophilin A was identified, which is involved in HIV and HCV-host interactions. Additionally, we extracted the name CPI-431-32 of cyclic peptide, inhibiting cyclophilin A [4]. Pharmaceutical agents were verified based on experimental data available in literature.

The presented approach allowed us to automatically process significant literature corpus on HIV and HCV mechanisms of pathological action and ways of their regulation. Our approach can be used in two main directions: (1) to automatically extract relation between proteins of various viruses and human and (2) to identify antiviral compounds molecular mechanisms of action including interactions with both viruses and human proteins. In addition, the developed workflow can be used in semi-automatic text corpora collection and annotations.

This work was supported by Russian Science Foundation Grant No. 19-75-10097.

1. Li, B., Dong, X., Zhang, W. et al. (2021) *IJMS*, 22, 9308.
2. SigNER. URL: http://www.way2drug.com/signer/
3. Naïve-Bayes CNER: http://www.way2drug.com/cner/
4. Gallay, P., Bobardt, M., Chatterji, U. et al. (2015) *PLoS ONE*, 10, e0134707.

COMPUTATIONAL APPROACH FOR IMPROVING OF KNOWN PERSPECTIVE SARS-COV-2 M^{PRO} INHIBITORS

***A.A. Bulygin*[1,2], *O.S. Fedorova*[1], *N.A. Kuznetsov*[1,2]**

[1]*Institute of Chemical Biology and Fundamental Medicine, Novosibirsk, Russia*
[2]*Department of Natural Sciences, Novosibirsk State University, Novosibirsk, Russia*

The COVID-19 pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) still poses serious threat to every man in the world. More and more variants of the virus emerge including ones with increased resistibility for developed vaccines. Thus, developing antiviral therapeutics is a high priority for the treatment of COVID-19. The main protease (Mpro) of SARS-CoV-2 is one of the good targets for hitting. This enzyme plays a central role in viral replication – cleavage of precursor polyproteins into functional viral proteins. Pfizer has developed two most known potent inhibitors of Mpro: PF-00835231 and PF-07321332. Both were tested in clinical trials [1]. PF-231 showed excellent inhibition ability, tolerable selectivity and unacceptable pharmacokinetic properties. The latter is mostly a result of weak solubility, which can be fixed by an addition of a highly hydrophilic group, which in its turn out the possibility of oral administration. PF-332 is an oral drug based on a different kind of a backbone. But its pharmacokinetic properties are still only acceptable with decreased selectivity and inhibition ability.

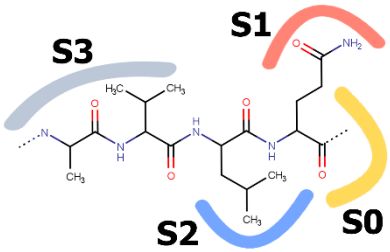


Figure 1. Mpro substrate.

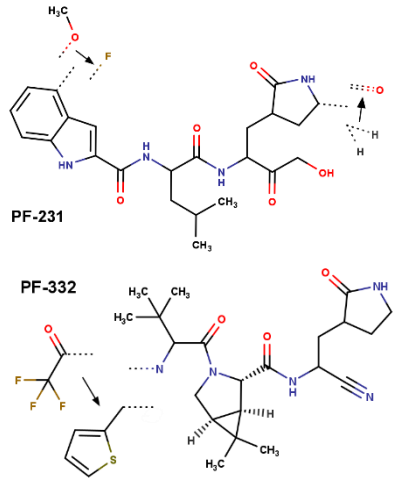


Figure 2. Changes in PFs.

In this work we tried to improve two most perspective inhibitors described above. To predict pharmacokinetic properties of new and known compounds SwissADME online tool was used [2]. Molecular dynamics simulations were conducted using GROMACS software [3] to obtain the structures of complexes of Mpro with known and new small molecules. To assess the strength of inhibitor-protein interaction binding energies were calculated by the MM/PBSA method [4].

First, we conduct an extensive review of *in vitro* studies of different inhibitors to define any correlations between structure and inhibition ability. Based on this, we can suggest four rules to construct a good inhibitor (Fig. 1): 1) the warhead (S0) for covalent binding should not be bulky; 2) S1 residue should resemble glutamine with one H-bond donor and one H-bond acceptor; 3) S2 residue should have 3-bond length and 4) a backbone need to be rigid for rapid binding and bulky somewhere at the N-end (S3) for selectivity.

Using these rules we found that the solubility of PF-231 can be increased without a loss in binding energy by two alterations (Fig. 2): indole's methoxyl can be replaced by a fluoride and the pyrrolidone residue can be replaced by a succinimide one (Fig. 2). This improves binding energy from -141±13 to -158±14 kJ/mol and average logarithm of solubility from -3.7 to -3.45. To increase binding ability of PF-332 the left side of the compound has to be changed. The best result gave a 5-member aromatic ring linked with the next residue through an aliphatic carbon (Fig. 2). Binding energy was improved from -132±13 to -180±15 kJ/mol and average logarithm of solubility from -4.0 to -3.9. Thus, there are ways to improve known inhibitors, which can be implemented in practice after synthetic chemistry optimization.

This study was supported by Russian-State-funded project No. 121112900214-2.

1. Owen, D.R., Allerton, C.M.N., Anderson, A.S. et al. (2021) *Science*, 374, 6575.
2. Daina, A., Michielin, O., Zoete, V. (2017) *Scientific Reports*, 7, 42717.
3. Abraham, M.J., Murtola, T., Schulz, R. et al. (2015) *SoftwareX*, 1-2.
4. Kumari R. et al. (2014) *Journal of Chemical Information and Modeling*, 54, 7.

TOWARDS THE *DE NOVO* DESIGN OF HIV-1 PROTEASE INHIBITORS BASED ON NATURAL PRODUCTS

A.L. Chávez-Hernández, K. Eurídice Juárez-Mercado, F.I. Saldívar-González, J.L. Medina-Franco

DIFACQUIM Research Group, Department of Pharmacy, School of Chemistry, National Autonomous University of Mexico

The acquired immunodeficiency syndrome (AIDS) caused by the human immunodeficiency virus (HIV) continues to be a public health problem. In 2020, 680000 people died from HIV-related causes, and 1.5 million people were infected. Antiretrovirals are only a way to control HIV infection but not to cure AIDS. As such, effective treatment must be developed to control AIDS. However, developing a drug is not an easy task, and there is an enormous amount of work and economic resources invested. For this reason, it is highly convenient to employ computer-aided drug design methods which can help generate and identify novel molecules. Using the *de novo* design, novel molecules can be developed. The use of natural product fragments as building blocks improves the diversity of compounds because they are characterized by a large diversity of functional groups, many sp^3 atoms, and chiral centers. Pseudo-natural products are a combination of natural products fragments that keep the desired structural characteristics from different natural products.

Bevirimat is a compound derived from betulinic acid and the first-generation maturation inhibitor. Bevirimat target the Gag polyprotein, inhibiting the action of HIV protease at its last cleavage event between capsid (CA) protein p24 and SP1. However, due to naturally occurring polymorphisms in Gag, 50% of patients exhibited reduced viral susceptibility to bevirimat, thus limiting its potential clinical utility [1]. Based on the structure-activity relationship (SAR) analysis developed by Zhao et al [2], the potency of bevirimat could be improved with residues from piperazine, 1,3-diaminoethane, and 1,3-diaminopropane attached to carbon17 of bevirimat. In this work, we develop a virtual focused compound library of HIV-1 viral protease inhibitors from natural product fragments using a *de novo* design based on natural product fragments and bevirimat as a template [3]. 1,534 compounds were developed using natural product fragments derived from the COlleCtion of Open NatUral producTs (COCONUT), the currently largest accessible database of natural products with more than 400,00 non-redundant compounds [4]. The 251 out of 1,534 compounds generated from COCONUT fragments, had physicochemical properties similar to FDA-Approved HIV-1 protease inhibitors and were estimated as easy synthesizable with the algorithm developed by Ertl and Schuffenhauer [5]. Compounds generated from COCONUT fragments were more diverse than compounds generated from two commercially available natural products: Enamine and ChemDiv. The estimation of ADME/Tox profiling showed that these compounds had adsorption and distribution like FDA-Approved HIV protease inhibitors. The excretion was different from FDA-Approved HIV-1 inhibitors but like FDA-Approved drugs.

The protocol presented in this work is general and can be used to build chemical compounds like bevirimat or other maturation inhibitors of HIV-protease. The interactive version of chemical space visualization of virtual compounds focused on HIV-1 viral protease inhibitors from natural product fragments is freely available at https://figshare.com/s/ceb58d58e8f5585ce67e.

1. Lazerwith, S.E. et. al. (2017). 5.19—New Antiretrovirals for HIV and Antivirals for HBV. Elsevier: Oxford, UK. pp. 628–664, ISBN 978-0-12-803201-5.
2. Zhao, Y.; Chen, C.-H., Morris-Natschke, S.L., Lee, K.-H. (2021).*Eur. J. Med. Chem.* 215, 113287.
3. Chávez-Hernández A.L., Juárez-Mercado K. E., Saldívar-González F.I., Medina-Franco J.L. (2021) *Biomolecules*, 11, 1805.
4. Sorokina, M., Steinbeck, C. *J.* (2020) *J. Cheminform.* 12, 20.
5. Ertl, P., Schuffenhauer, A. (2009). *J. Cheminform*, 1, 8.

DIFFERENTIAL EXPRESSION OF ALTERNATIVE SPLICE-FORMS IN CANCER

G.V. Dolgalev, E.V. Poverennaya

Institute of Biomedical Chemistry, Moscow, Russia

The cancer is a heterogenous disease in which multiple pathways can be hijacked to give rise to transformed cells. One of the mainstay molecular pathways affected in cancer is an alternative splicing. Dysregulation of alternative splicing can lead to expression of transcript isoforms that encode protein products with cancer-relevant functions [1]. Accordingly, understanding functional capacity of transcript isoforms enriched in particular cancer types can potentially reveal the molecular basis of malignant transformation. In this study, RNA sequencing data from The Cancer Genome Atlas (TCGA) project has been analyzed to identify genes which produce transcript isoforms with different pattern of expression between malignant and normal tissue samples.

To this end, transcript expression data of 23 cancer types and corresponding normal tissues from TCGA (total of 8286 samples) have been downloaded and transcripts were mapped to UniProt database in order to focus on curated protein-coding isoforms only. Next, for each gene and tissue type, an existence of prevailing transcript isoform (one that has the highest expression levels in more than 80% of samples) has been analyzed. After comparing prevailing isoforms between malignant and normal samples, 695 instances of difference in prevailing transcript isoforms were detected. Furthermore, 28 genes have been found to differentially express spliceforms in the same way in more than three cancer types.

To explore the potential relevance of identified cases, corresponding expression data from other large RNA-Seq databases such as GTEx, Human Protein Atlas (HPA) and DepMap project have been compared to the obtained results. Strikingly, most of the identified cases demonstrated expression patterns unique only to the TCGA data, which highlighted their potential relevance to the process of malignant transformation.

This study was supported by RSF No. 21-74-10061.

1. Y. Zhang et al. (2021) *Sig. Transduct. Target. Ther.*, 6, 78.

TESTING THE ACTIVITY OF BIGUANIDES AND SOME NOVEL DESIGNED MOLECULES AGAINST SARS-COV-2 PROTEINS, IN SILICO STUDY

***M. Efendi*[1], *T. Taskin-Tok*[1,2]**

[1] *Gaziantep University, Faculty of Arts and Sciences, Department of Chemistry, Gaziantep, Turkey*
[2] *Gaziantep University, Institute of Health Sciences, Department of Bioinformatics and Computational Biology, Gaziantep, Turkey*

COVID-19 continues to threaten lives of thousands of people around the world, the main pathogen of this disease is severe acute respiratory syndrome coronavirus 2 or as known (SARS- CoV-2). After hijacking human cell, this virus gets inside it, then the replication and copying processes are initiated to form new copies of this virus, which attack new cell, and so the infection occurs. There are many structural and non- structural proteins responsible for this replication process including; main protease (MPro), papain like protease (PLPro), RNA dependent RNA polymerase (RdRp). Through our study, we aim to inhibit roles of these proteins, by applying molecular docking to investigate the effectiveness of biguanide derivatives (Cycloguanil, Phenformin, Proguanil, Moroxydine, Buformin and Metformin) against these targets, and then molecular dynamic simulation method is applied to investigate the stability of complexes of best docking results. Also, we, in silico, modify biguanide derivatives to get better results, as a last step, we apply Absorption, Distribution, Metabolism, Excretion and Toxicity test (ADMET) to those modified compounds which showed good docking results. As a result of our study, phenformin played an inhibition role against RdRp and MPro proteins, proguanil was recommended to be used against PLPro, at the same time, 10 of our modified compounds gave a promise results against this three targets, so other drug development steps could be applied in order to investigate this activity.

STUDY OF GENES ASSOCIATED WITH THE DEVELOPMENT OF CROHN'S DISEASE AND RECONSTRUCTION OF GENE NETWORK USING BIOINFORMATICS TOOLS

***A.D. Efremova*[1], *Y.L. Orlov*[1,2]**

[1]*I.M. Sechenov First Moscow State Medical University of the Russian Ministry of Health (Sechenov University), Moscow, Russia*
[2]*Novosibirsk State University, Novosibirsk, Russia.*

Genomic association studies and computerized meta-analyses revealed and confirmed a set of loci of susceptibility to Crohn's disease on human chromosomes [1], [2]. Understanding the role of the innate immune response in Crohn's disease was the result of the discovery of the corresponding susceptibility loci in genetic studies. However, although improved statistical models and increased sample sizes from consortia will increase the number of susceptibility loci identified in the future, this component still explains only slightly more than 20% of the heritability of Crohn's disease, which underlines the importance of studying genetic, epigenetic and environmental factors [1]. Our work is aimed at building (collecting) a list of genes associated with the development of Crohn's disease, analyzing the categories of gene ontologies for such a list and reconstructing the gene network. For the key genes of the disease obtained by analyzing the structure of the gene network, options for the search for drugs (substances interacting with this protein) are considered.

To build a list of genes associated with hereditary predisposition to Crohn's disease and to analyze the genes of Mendelian inheritance, the online resource OMIM (Online Mendelian Inheritance in Man, https://omim.org) was used [3]. To search and analyze significant categories of gene ontologies for this group of genes, the resulting list of human genes was downloaded via the DAVID interface (Database for Annotation, Visualization and Integrated Discovery) (https://david.ncifcrf.gov/summary.jsp) and PANTHER [4]. The categories of gene ontologies for molecular functions were also calculated. GeneMANIA (https://genemania.org/) and STRING-DB (https://string-db.org/) tools were used to reconstruct the gene network of interactions of genes associated with Crohn's disease.

The GO analysis revealed that the most significant categories for genes associated with Crohn's disease are signaling pathways mediated by tumor necrosis factor and caused by cytokines, as well as reactions to the drug and cytosolic processes. Due to the identification of susceptibility loci, it is possible to conclude about violations of immune regulation in the intestine, as well as the presence of categories of cytosolic processes and signaling pathways allows us to determine the relationship of the disease with intracellular inflammation. DAVID and PANTHER for Crohn's disease genes confirm the categories of gene ontologies of protein binding, membranes and intracellular processes. The most significant categories of gene ontologies for molecular functions are the categories of protein binding, binding of enzymes and cytokines, and other binding of protein-containing complexes. The resulting gene network is quite coherent, although the connections were exposed only by the parameters protein-protein interactions and co-expression. Overall, search for novel drug targets in Crohn's disease remains challenging problem [5].

1. Baumgart, D.C., Sandborn, W.J. (2012) *Lancet*, 380(9853), 1590-605.
2. Khor, B., Gardet, A., Xavier, R.J. (2011) *Nature*, 474(7351), 307-17.
3. Amberger, J.S., Bocchini, C.A., Scott, A.F., Hamosh, A. (2019) *Nucleic Acids Res.* 47(D1), D1038-D1043.
4. Mi, H., Muruganujanm A, Thomasm PD. (2013) *Nucleic Acids Res.* 41, D377-86.
5. Antunes, J.C., Seabra, C.L., Domingues, J.M. et al. (2021) *Nanomaterials (Basel)*, 11(8), 2035.

ACTIVITY PREDICTION OF SARS-COV-2 MPRO INHIBITORS BASED ON ENSEMBLE DOCKING AND MACHINE LEARNING

***A.D. Fomina*[1,2], *D.I. Osolodkin*[2]**

[1]*Department of Chemistry, Lomonosov Moscow State University, Moscow, Russia*
[2]*FSASI "Chumakov FSC R&D IBP RAS", Moscow, Russia*

The COVID-19 pandemic has been going on for two years now. During this time, vaccines and approved drugs for SARS-CoV-2 have been developed, but the design of more selective and effective antiviral drugs remains an urgent problem. One of the most promising targets for the anticoronavirus drugs design is the main protease (Mpro) [1]. The amount of available structural data allows the most efficient use of the ensemble docking method, which uses the results of docking into various structures to rank ligands for virtual screening.

The aim of our work was to develop a virtual screening method for SARS-CoV-2 Mpro protease inhibitors based on ensemble docking and machine learning. The machine learning model is used to improve generated during the docking process poses ranking.

As a training set, a library of 6897 compounds with experimentally determined percentage of SARS-CoV-2 protease Mpro inhibition at a concentration of 20 μM was used. Compounds with inhibition percentage of more than 50% were assigned active. 70% of the compounds were assigned into training set, and the remaining 30% were split in half into test and validation sets. The ratio of active to inactive compounds in the training, test and validation sets were 0.034, 0.034, 0.034 respectively.[2].

In January 2022, 415 SARS-CoV-2 Mpro structures were available in the PDB database. The ensemble of 3CLpro protease structures was composed of mature, non-oxidized, fully resolved structures, which have the highest all-atom pairwise RMSD between the conformations of the active site residues. Six structures were selected to establish the final ensemble [3].

The training set was comprised of the structures of the best protein-ligand complexes obtained by docking compounds with known activity into the structures of the ensemble using DOCK6.9. The vector description of the complex was made using interaction fingerprints characterizing the type of contact between the ligand atoms and the nearest protein atoms. Fingerprints were calculated using the Flare 5.0.0 (Cresset, UK).

Various machine learning models (random forest, gradient boosting, support vector machine, deep learning) were trained on vector descriptions of ligand-protein complexes and percent inhibition of Mpro protease by the ligand to classify active and inactive molecules. The best by AUC and ROC-curves result was shown by the random forest model with AUC of 0.79 on validation dataset.

This study was supported by the Non-commercial Foundation for the Advancement of Science and Education INTELLECT and State Research Funding No. FNZG-2022-0002.

We thank Cresset team for Flare academic license.

1. Zakharova ,M.Yu. et al. (2021) *Frontiers in Pharmacology* 2021(12), 773198.
2. Kuzikov M. et al. (2021) *ACS Pharmacology & Translational Science*, 3(4), 1096–1110.
3. Fomina, A.D., Osolodkin, D.I. (2021) BCADD, Moscow, 70.

STRUCTURAL EVALUATION OF N-BENZYL-1-[1-(NAPHTHALEN-1-YL)ETHYL]PIPERIDINE-4-CARBOXAMIDE DERIVATIVES IN TARGETING SARS-COV-2 PAPAIN-LIKE PROTEASE: AN *IN-SILICO* PERSPECTIVE

D.M.D. Albar, K.M.D. Gatdula, J.R.R. Javier, C.C. Deocaris

Department of Physical Sciences, College of Science, Polytechnic University of the Philippines, Manila, Philippines

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the causative agent for COVID-19, is a novel human betacoronavirus that resulted in a serious global pandemic. The papain-like protease (PLpro) is one of the proteases produced by SARS-CoV-2. The enzyme participates in viral polyprotein cleavage, deubiquitination and deISGylation. The role of PLpro in viral replication makes it a good pharmaceutical target. Several inhibitors have been studied for every enzyme encoded by the coronavirus genomes. The sequence, structure, and functional conservation of PLpro among all the coronaviruses suggest that therapeutics targeting SARS-CoV-1 PLpro may also be effective towards PLpro of SARS-CoV-2. An approach to find PLpro inhibitors for COVID-19 is to evaluate previously identified inhibitors from similar enzymes**.** The molecule N-(1,3-benzodioxol-5-yl methyl)-1-[(1R)-1-naphthalen-1-yl ethyl] piperidine-4-carboxamide (GRL-0667) has been identified as a competitive naphthalene-based inhibitor of SARS-CoV PLpro. The derivatives of this molecule have also shown inhibitory properties on SARS-CoV-1 PLpro. In this study, the molecule GRL-0667 and its derivatives were evaluated to determine its physicochemical, and potential pharmacodynamics and pharmacokinetic parameters against the PLpro of SARS-CoV-2. Bioinformatic tools such as: Molinspiration, OSIRIS DataWarrior, admetSAR and AutoDock Vina was employed to evaluate GRL-0667 and its derivatives as potential ligands within the substrate-binding site of the SARS-CoV-2 PLpro macromolecule.

The Molinspiration simulations revealed that some GRL-0667 derivatives failed Lipinski's rule (rule of 5) making these molecules unviable for further pharmacologic studies. On the other hand, all GRL-0667 derivatives had good oral bioavailability based on Veber's rule. The bioactivity scores indicated that the molecules were biologically active as protease inhibitor and G protein-coupled receptor ligand. The molecules tested showed a potential tumorigenic toxicity. But then, these are safe in terms of reproductive and irritant effects. The absorption, distribution, metabolism, excretion, and toxicity (ADMET) profiles of the molecules suggest its permeation through the blood-brain barrier (BBB) and human intestine absorption (HIA). Docking simulation with the SARS-CoV-2 PLpro (PDB ID: 6WUU) showed binding energies in the range of -8.24 to -9.61 kcal/mol by the GRL-0667 and its derivative. A strong binding energy of -9.61 kcal/mol was noted for the ligand 3i, a GRL-0677 derivative having a 3,4-difluoro substituent in its R3 location. The results of this study may present an opportunity for the development of antiviral drugs targeting SARS-CoV-2.

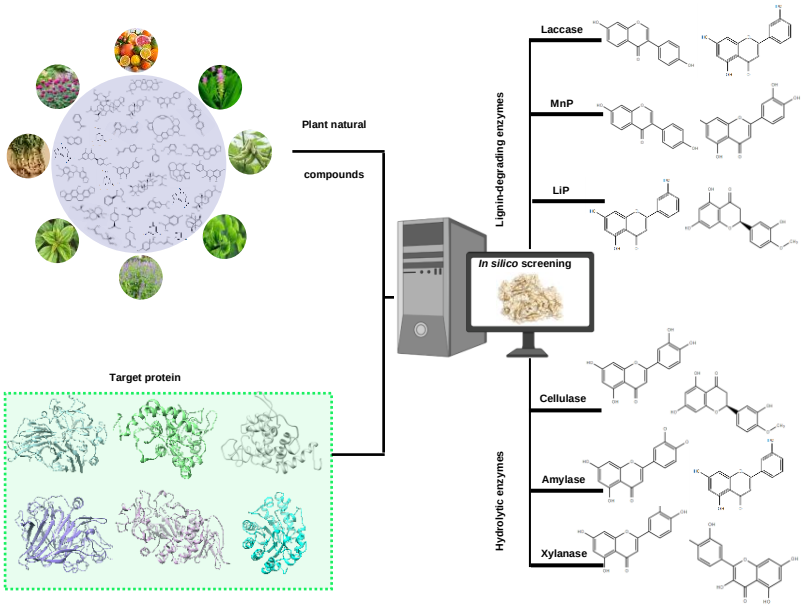
VIRTUAL SCREENING OF PLANT-DERIVED COMPOUNDS TARGETING HYDROLYTIC AND LIGNIN DEGRADING ENZYMES OF *GANODERMA BONINENSE*

***H. Hadiatullah*[1], *R.A. Putranto*[2], *G.W. Permatasari*[2], *R.T. Saptari*[2], *Z. Yuchi*[1]**

[1]*Tianjin Key Laboratory for Modern Drug Delivery & High-Efficiency, Collaborative Innovation Center of Chemical Science and Engineering, School of Pharmaceutical Science and Technology, Tianjin University, Tianjin, China.*

[2]*Indonesian Research Institute of Biotechnology and Bioindustry, Bogor, Indonesia*

Ganoderma boninense is a primary pathogen of oil palm (*Elaeis guineensis*) known as basal stem rot disease (BSR). *G. boninense* invades the oil palm by first degrading the lignin employing lignolytic enzymes (laccase, manganese peroxidase (MnP), and lignin peroxidase (LiP)), and then degrading the cellulose with hydrolytic enzymes (cellulase, amylase, and xylanase) [1]. In the present high-throughput virtual screening study, we used *in silico* techniques like receptor-ligand docking, physicochemical characteristics, and toxicity prediction to screen plant-derived substances potentially having inhibitory effects against *G.boninense*. Following a comprehensive screening analysis, three novel multi-target natural compounds, including apigenin, luteolin, and hesperetin, would inhibit the activity of lignin-degrading (laccase, manganese peroxidase (MnP), lignin peroxidase (LiP)) and hydrolytic (cellulase, amylase, xylanase) enzymes of *G. boninense*. These three compounds have a higher binding affinity against the target protein and are nontoxic. Additionally, our findings shed light on the mechanism of action of these compounds and suggested that their structural diversity may contribute to the development of novel anti-*G. boninense inhibitors using* structure-based optimization strategies. Concerning the utility of plant-derived compounds in the formulation of a variety of therapies, including those in pharmaceutical and agricultural fields, we propose that these compounds serve as excellent lead candidates for the development of potential *G. boninense*-inhibtors, pending *in vitro* and *in vivo* validation.



This study was supported by Ministry of State-Owned Enterprises of Indonesia and National Natural Science Foundation of China (No. 32022073 and 31972287)

1. Paterson, R.R.M.J.M. (2019) *Microorganism*, 7, 24.

STRUCTURAL OPTIMIZATION OF TUBULIN INHIBITORS

***A. Ivanova*, O. Mokshyna, P. Polishchuk**

Institute of Molecular and Translational Medicine, Faculty of Medicine and Dentistry, Palacký University and University Hospital in Olomouc, Olomouc, Czech Republic

Tubulin inhibitors prevent microtubule formation and mitosis progression making them useful for anticancer therapy. Earlier in our institution were synthesized and investigated a set of tubulin inhibitors bearing a novel scaffold. The lead compounds have high activity and selectivity but unfavorable physicochemical properties, in particular high lipophilicity, making them hard to develop further. The goals of our work were: i) study the structure-activity relationship of highly active and selective tubulin inhibitors previously synthesized in our institute; ii) establish their binding mode and suggest possible directions of modifications, iii) design new analogs with improved physicochemical properties.

Docking of previously synthesized compounds to different conformations of tubulin allowed to establish that the protein structure from the complex with combrestatine-A4 (PDB: 5LYJ) is more suitable and result in higher correlation of activity with calculated docking scores than docking to 4O2B and 5CA1 tubulin structures. This was additionally confirmed by 100 ns molecular dynamic (MD) simulations which demonstrated high stability of the established pose of the lead compound. MD study allowed to establish that the majority of protein-ligand contacts have hydrophobic nature, but it was found that a nitrogen in the core part of the lead molecule can create a hydrogen bond through a water bridge and this contact persists in course of the simulation. Based on the established binding pose we suggested promising directions of modification of the lead compound. To design new compounds we preserved important scaffold features and enumerated possible analogs by CReM tool, which main feature is generation of synthetically feasible structures [1]. Finally compounds with desired physicochemical properties, in particular low lipophilicity, were selected and evaluated by docking procedure and the most promising ones were suggested for synthesis and biological experiments. Some of the suggested analogs have been already synthesized and tested. They demonstrated the same level of activity/selectivity as the lead compound and improved lipophilicity. Although there is still a significant part of molecules in the queue for synthesis and experimental validation.

1. Polishchuk, P. (2020) *J Cheminform.*, 12, 28.

IN SILICO ANALYSIS AND DISCOVERY OF POTENTIAL DRUG LIGANDS OF HYPOTHETICAL PROTEINS OF *TALAROMYCES MARNEFFEI*

***A. Juanga*, P. Sevilla, J.K. Velasco**

Department of Biology, College of Science, Polytechnic University of the Philippines, Manila, Philippines

The increase in the emergence of infectious diseases causes a high demand to develop new treatments and discover new drugs to combat such diseases. *Talaromyces marneffei* is a saprozoonotic fungus endemic in Southeast Asia which can be transmitted from animal to humans and causes infections, especially to the immunocompromised individuals that could lead to death. In this study, we used hypothetical proteins to come up with a potential target for drug discovery against *T. marneffei*. As there were no current studies with a similar approach to us, our findings could open new ways and opportunities to design drug molecules against this deadly disease. We also focused on finding nutraceutical drugs instead of pharmaceutical drugs because of their advantages such as they are accessible and rarely have side effects. In this study, we collected 9,673 hypothetical proteins of *T. marneffei* (ATCC 18224, PM1, and 11CN-03-130) from UniProt database. With the use of *in silico* approach, the hypothetical proteins went to a series of screenings under the following bioinformatics tools and software: hmmscan, Blast2GO, KEGG pathway, DrugBank, SWISS-MODEL, PROCHECK, ProSA-web, and STRINGdb. In selecting the best drug target among the hypothetical proteins, we used the following criteria: (1) it has a good homology model, (2) it has a great number of PPIN and has an interesting COG, and (3) it has a great number of interacting drugs. We finally came up with A0A7C8H5I4 protein and characterized it further into physiological and physicochemical properties using MEGA-X, InterPro, NCBI CDD, ProtParam tool, NetSurfP-2.0, and WoLFPSORT. The study shows 14-alpha sterol demethylase and Eburicol 14-alpha demethylase as the paralogs of A0A7C8H5I4 protein which were known to be drug targets of the current azole antifungals. In the end, we generated five nutraceutical drugs using DrugBank but only four—niacin, melatonin, ademetionine, and menadione—had small K_D values or possessed great binding affinity according to the molecular docking experiments using AutoDock Vina. Our study is significant in drug discovery and the emerging drug resistance of *T. marneffei*.

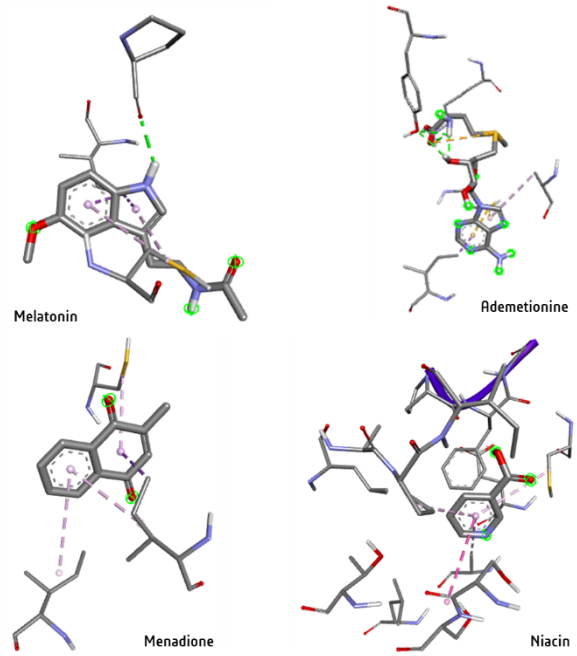


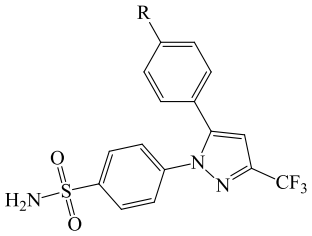
Figure 1. AutoDock Vina receptor-ligand visualization.

NEW INHIBITORS OF CYCLOOXYGENASE ISOFORMS IN A NUMBER OF SOME PYRIMIDINE DERIVATIVES

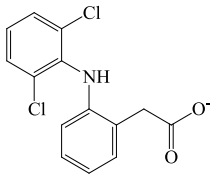
***Y.V. Kiseleva*[1], *V.R. Khairullina*[2], *Y.Z. Martynova*[1], *A.G. Mustafin*[1,2]**

[1]*Department of Chemistry, Bashkir State University, Ufa, Russia*
[2]*Ufa Institute of Chemistry of the Russian Academy of Sciences, Ufa, Russia*

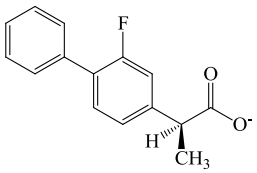
By the method of molecular docking using the AutoDock 4.2.6 program, the steric complementarity of 17 uracil derivatives with cyclic and acyclic sulfur- and oxygen-containing substituents in the pyrimidine cycle, with active centers of isoforms of cyclooxygenases (COX) was studied. As a model of COX-1 and COX-2 proteins, macromolecules with codes B and A were chosen.3n8x and 1pxx in the PDB protein databank, respectively). The protein molecules in the calculations were rigid, while the ligand molecules were mobile. The size of the three-dimensional box, in which molecular docking of ligands was carried out, in all cases was 50 steps with a step marking of 0.375 Å. The position of the natural substrate of these enzymes - arachidonic acid and the structures of reference inhibitors of COX isoforms, which are the active components of nonsteroidal anti-inflammatory drugs (NSAIDs) Celecoxib, Flurbiprofen, Diclofenac (Fig. 1), was taken as the center of the box.



R=CH_3 DV NSAIDs "Celecoxib";
R=Br leader connection
SC-558



DV NSAIDs
«Diclofenac»



DV NSAIDs
Flurbiprofen

Figure 1. Structures of reference inhibitors of COX isoforms.

Based on the results of computational experiments, 2 leading compounds (16 and 17) were selected for synthesis, which theoretically can be effective inhibitors of both forms of COX, and therefore detect a pronounced anti-inflammatory effect in vivo. These compounds are conjugates of 5-hydroxy-1,3,6-trimethyluracil with N-phthalyl-L-amino acids, respectively. A comparative analysis of the free binding energies calculated using the AutoDock 4.2 scoring function suggested that both of these compounds would be characterized by increased selectivity of action against COX-2. These compounds are promising for further research as potential anti-inflammatory drugs.

The reported study was funded by the Russian Science Foundation for Basic Research according to the research project No. 19-73-20073.

1. URL: https://www.rcsb.org/
2. URL: http://vina.scripps.edu/

DIMERIC STATES OF TRANSMEMBRANE SEGMENTS OF THE DDR1 RECEPTOR PREDICTED BY ATOMISTIC MODELING

***E. Kozlov*[1], *A. Kuznetsov*[1,2], *R. Efremov*[1,2]**

[1]*Moscow Institute of Physics and Technology (National Research University), Moscow, Russia*
[2]*Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry of the Russian Academy of Sciences, Russia*

Discoidin domain receptor 1 (DDR1) is a membrane protein from the receptor tyrosine kinases (RTKs) family, activated by collagen and involved in cell growth, development, differentiation, and proliferation. DDR1 dysfunction can lead to the development of fibrosis, arthritis, and cancer. DDR1 consists of three structural domains: extracellular, responsible for binding to collagen; intracellular, providing phosphorylation of tyrosine residues; and transmembrane (TM), facilitating signal transmission across the membrane [1, 2]. All of these domains are involved in the formation of dimeric and oligomeric forms of DDR1. The spatial structure of the full-size DDR1 is unknown, since there is no available data on the TM dimer. This makes it difficult to study the mechanism of receptor activation and search for effective methods to modulate its activity. Based on literature data it was suggested that the activation of the DDR1 receptor can occur through heterodimerization with other RTKs and/or when the membrane environment changes [1, 3].

This work is aimed at studying the structure of the DDR1 TM dimer, identifying key amino acid residues that stabilize the active/inactive states. Hopefully, the results will help in the rational design of interceptor peptides capable of binding DDR1 TM domains, thus modulating its activity in the cell membrane.

So, based on amino acid sequence and molecular hydrophobicity maps we compared TM domain of DDR1 with TM domains of other RTKs. At the same time, we used PredDimer web service (model.nmr.ru/preddimer) to build preliminary *de novo* structures of possible TM dimers. The latter were clustered and compared with homology-based models. Stability of the resulting models in the membrane environment was assessed via molecular dynamics simulations implemented in the GROMACS software. For all stable dimeric structures, the free energy of dimerization was evaluated, and the most probable conformations of TM dimers were selected.

Comparing amino acid sequences and surface molecular hydrophobicity among TM segments of RTKs, DDR1 was found to be close to the ErbB2 receptor, for which two different experimental structures exist. 9 *de novo* structures were built using PredDimer, with two of them being significantly similar with structures of active and inactive ErbB2 TM dimer, respectively. After stability and free energy evaluation, 5 stable models of DDR1 TM dimer were selected, and two homologous with ErbB2 among them. Characteristic dimerization interfaces were identified, suggesting two possible ways of formation of DDR1 oligomeric forms using a hydrophilic glycophorin-like interface or an extended hydrophobic pattern. This allows us to propose the existence of an ensemble of different states of TM segments corresponding to the active and inactive states of the DDR1 receptor.

Having DDR1 TM domain being similar to ErbB2, an assumption was made about apparently similar activation mechanisms and/or heterodimerization in membranes. Moreover, interceptor peptides and mutations have been previously described for ErbB2, that can also be tested for DDR1 TM domain and help in development of new strategies of DDR1 activity modulation.

This work was supported by RSF grant No. 18-14-00375.

1. Carafoli, F., Hohenester, E. (2013) *Biochimica et Biophysica Acta* (*BBA) - Proteins and Proteomics*, 1834, 2187–2194.
2. Mihai, C., Chotani, M., Elton, T.S., Agarwal, G. (2009) *Journal of Molecular Biology*, 385, 432–445.
3. Agarwal, G., Smith, A.W., Jones, B. (2019) *Biochimica et Biophysica Acta (BBA) – Molecular Cell Research*, 1866(11), 118496.

$BLT_{1/2}$ AGONIST BINGING MODEL EXPLAINS NEW EXPERIMENTAL DATA ON LIGAND BINDING

G. Kurakin

Pirogov Russian National Research Medical University, Moscow, Russia

In 2020, we have published a paper on comparative computational modelling on agonist binding to the leukotriene receptors BLT_1 and BLT_2 [1]. This paper was presented and the Symposium on Bioinformatics and Computer-Aided Drug Discovery in 2020.

In that paper, we proposed that Arg156 BLT_1 (Arg160 BLT_2) forms a salt bridge with the carboxylate head of ligand which is crucial for binding, while Glu185, Tyr237 BLT_1 (Tyr240 BLT_2), and Asn241 BLT_1 (Asn244 BLT_2) form a network of hydrogen bonds with a 'tail' hydroxyl of a ligand. In our model, ligand specificity was provided by His94 BLT_1 (Tyr98 BLT_2) determines the hydrophilic or hydrophobic site near the 5' position of the ligand. At the moment of publishing, our model explained available experimental data on BLT_1 and BLT_2 to a maximal extent. However, since the moment of publishing this work, some other papers on $BLT_{1/2}$ ligand binding were published. So, this report is to re-evaluate our model in terms of consistency with experimental data and other models.

First of all, other models published since 2020 differ significantly from our model. In one case, experimental data could falsify (in Carl Popper's meaning) our results. It is the case of the paper by Giusti et al. [2], where the NMR data on ligand conformers were obtained. One of the co-authors, Dr. Laurent Catoire, kindly provided us their NMR conformers for further investigations. We have revealed that these conformers cannot be fully superimpose to conformers predicted by us. However, they appeared to be able to form similar bonds and occupy similar position upon flexible docking with ROSIE. We have interpreted this result as a partial consistency with our prediction and a chance to unite these NMR data and other experimental results under our model.

In the rest two cases, new experimental data were consistent with our predictions even if the co-published models were not. For instance, Kim et al. [3] have firstly provided experimental evidence for the Tyr240 – Asn244 – "tail" OH hydrogen bond network despite they have proposed quite other model of ligand binding. So, our first prediction came true. Our second prediction came true with the publication of the paper by Michaelian et al. [4]: they confirmed that the ligand carboxyl group forms a salt bridge with Arg156 as we had predicted earlier. It is worth noticing that these X-ray data did not correspond with the models by Giusti et al. [2] and Kim et al. [3], but they did correspond with our model corrected with NMR data of Giusti et al. [2]. We have concluded that our model could be used for BLT-targeted drug design as a working model of agonist binding.

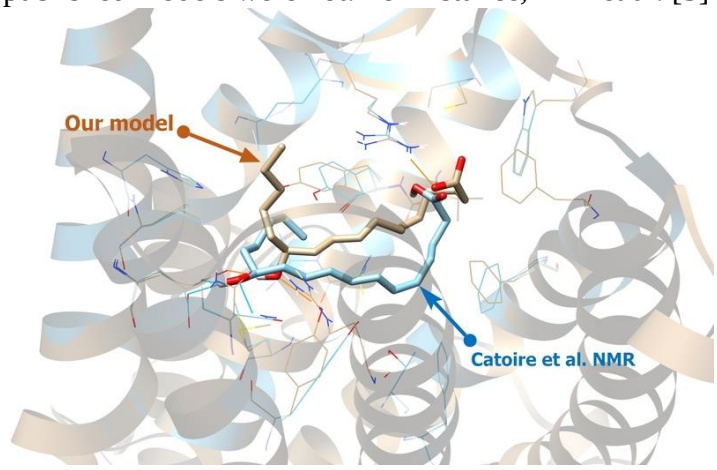


Figure 2. Comparison of predicted and NMR-derived conformers.

1. Kurakin, G.F., Lopina, N.P., Bordina, G.E. (2020) *Molecular Biology*, 54, 299–309.
2. Giusti, F., Casiraghi, M., Point E. et al. (2020) *Scientific Reports*, 10, 2630.
3. Kim, M., Wei, J.-D., Harmalkar D.S. et al. (2020) *ACS Med. Chem. Lett.*, 2020, 11, 8, 1529–1534.
4. Michaelian, N., Sadybekov, A, Besserer-Offroy E. et al. (2021) *Nature Communications*, 12, 2971.

MOLECULAR DYNAMIC PHARMACOPHORE AND ITS APPLICATION IN DESIGNING NOVEL MARK4 INHIBITORS

***A. Kutlushina*, O. Mokshyna, L. Hruba, S. Gurska, M. Hajduch, P.Dzubak, P. Polishchuk**

Institute of Molecular and Translation Medicine, Olomouc, Czech Republic

Microtubule affinity-regulating kinase 4 (MARK4) belongs to the family of serine/threonine kinases and is the mammalian homolog of partition-defective 1. Unlike other MARK isoforms, MARK4 is unique in its ability to exhibit association with microtubules through phosphorylation of MAP2, MAP4 and tau proteins. Microtubules are involved in many biological processes, thus targeting therapies against MARK4 could possibly provide new promising treatment against Alzheimer's disease, cancer, atherosclerosis and type II diabetes. In our days, existing active MARK4 inhibitors are not specific. This inspired us to search for MARK4 inhibitors with a new scaffold, which will allow us to expand the chemical space of active inhibitors of MARK4 and increase the chances of finding specific molecules to the target kinase.

Pharmacophores are widely used in drug design at the first stages of virtual screening, since they are abstract and allow finding molecules with new scaffolds. We developed and applied a workflow which includes virtual screening using pharmacophores retrieved from molecular dynamics (MD) trajectories and molecular docking. First, we selected 4 ligands from our previous studies which demonstrated high inhibitory activity against MARK4. These ligands were docked to MARK4 (PDB 5ES1) and afterwards 100 ns MD simulations of the ligand-protein complexes were performed. From each frame of these MD trajectories we extracted 3D pharmacophores by the previously developed pharMD tool [1]. To reduce redundancy, identical or very similar pharmacophores were identified and removed by comparing their 3D pharmacophore hashes, which were calculated by pmapper[2]. The remaining pharmacophore models were used to screen the Enamine database consisting of more than 2 million compounds. The compounds were ranked by their consensus scoring. The 1000 top scored compounds were docked to MARK4 using Autodock Vina. Based on their docking scores and poses, the best 24 compounds were selected and purchased from Enamine. Two of these compounds demonstrated activity in primary screening. Measurement of their IC_{50} values revealed moderate activity (IC_{50}=12 μM) for one of them and high activity (IC_{50}=30 nM) for the other. The activity of the latter compound was comparable to reference compounds used on the first stage of the study. It is noteworthy, that the scaffold of this newly identified inhibitor was completely different from those reference structures. We also measured activity of these active compounds to other MARK family kinases and found that the most active compound was highly selective over MARK1 and MARK2 but has a high activity against MARK3 (IC50=9 nM). Gabriel von Euler et al. showed in their article 2014 that MARK3 can be associated with Alzheimer's disease.

This work was funded by the INTER-EXCELLENCE LTARF18013 project (MEYS), the European Regional Development Fund - Project ENOCH (No. CZ.02.1.01/0.0/0.0/16_019/0000868) and ELIXIR CZ research infrastructure project (MEYS Grant No: LM2018131)

1. Polishchuk, P., Kutlushina, A. et al. (2019) *International Journal of Molecular Sciences*, 20, 5834.
2. Kutlushina, A., Khakimova, A., Madzhidov, T., Polishchuk, P. (2018) *Molecules*, 23, 3094.

AliNA – A DEEP LEARNING-BASED PROGRAM FOR PREDICTION OF RNA SECONDARY STRUCTURE WITHOUT SPECIFICATION OF THERMODYNAMIC PARAMETERS

S. Nasaev[1], A. Mukanov[2], I. Kuznetsov[3], A. Veselovsky[1]

[1]Institute of Biomedical Chemistry, Moscow, Russia
[2]A.M. Butlerov Institute of Chemistry,Kazan Federal University, Kazan, Russia
[3]Moscow University of Finance and Law, Moscow, Russia

Nowadays there are a number of discovered noncoding RNA variations participating in different cell processes. One of the required tasks in investigation of their functions and mechanism of influence on cells is prediction of RNA structures. The spatial structure is important for functioning of numerous types of RNA. It also provides possibility of ribosome interaction or RNA interference, etc. Alteration of RNA spatial structure is important for riboswitches action. The RNA spatial structure can also affect PCR or sequencing results. RNA structural formation is hierarchical and secondary structure determines the tertiary folding. The modern RNA secondary structure prediction methods are based on homology modelling and dynamic programming. Homology-based prediction requires a large number of different RNA structures and it is unusable for prediction of non-homologous sequences. The other methods are based on finding minimum energy using scoring functions, whereas the structure of a natural RNA often exists in local minimum. Pseudoknots can often be found in RNA structures and their structure prediction is a tremendous problem. The computational complexity of pseudoknot prediction can rise up to $O(N^6)$ (where N is the length of the nucleotide sequence). We present the AliNA program which is a secondary structure prediction method based on deep learning. A significant advantage of this method is usage of dinucleotides pairs as basic interacting units. This approach makes it possible to predict any structures including those containing pseudoknots with high precision.

The AliNA method shows better secondary structure prediction compared to other modern methods (SPOT-RNA, MXFold and DotKnot) regardless of the selection chosen for comparison. Despite the difficulty of pseudoknots prediction AliNA surpasses the rest of the methods including the DotKnot algorithm designed specifically for this goal. The method is free available.

The work was performed within the framework of the Program for Basic Research in the Russian Federation for a long-term period (2021-2030) (No. 122030100170-5).

PREDICTING POTENTIAL TARGETS OF RRx-001 AS A RADIOPROTECTOR: AN *IN SILICO* TARGET FISHING APPROACH

R.J. Co[1], L.M. Lacandula[*,1], D.A. Tina[*,1], L.V. Alvarez[1], C. Deocaris[2], V. Alinsug[3]

[1]Department of Biology, College of Science, Polytechnic University of the Philippines, Manila, Philippines

[2]Biomedical Research Section, Philippine Nuclear Research Institute, Department of Science and Technology, Quezon City, Philippines

[3]Atomic Research Division, Agricultural Research, Philippine Nuclear Research Institute, Department of Science and Technology, Quezon City, Philippines

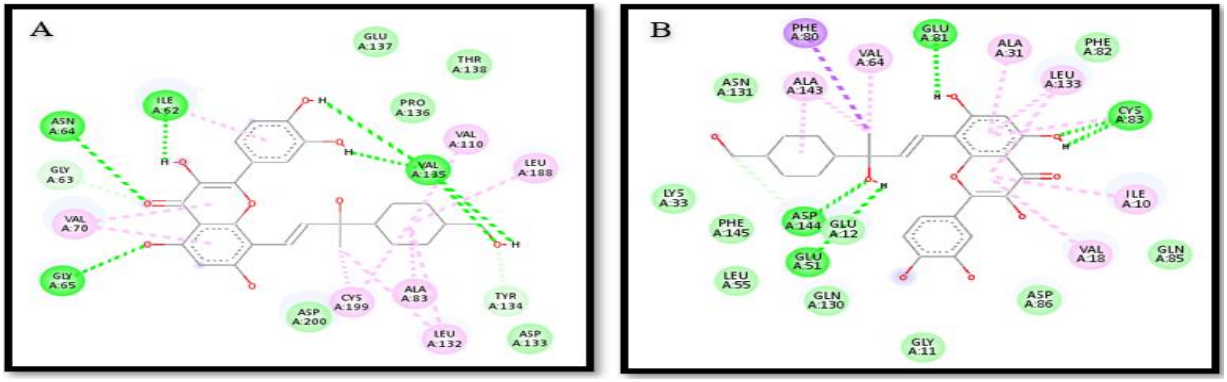
Radioprotectors are agents used to inhibit radiation-induced toxicities that cause cellular damage, which is critical in treating cancer and its derivatives. RRx-001, a small molecule sourced from space, is well known for its radiosensitizing activities. Several studies have also reported on its various cancer-related pharmacological properties through antiangiogenic, antioxidant, apoptotic, and nitric oxide pathways without observed dose-limiting toxicities. This paper aimed to predict and evaluate the potential targets of RRx-001 using bioinformatic tools as no specific computational target fishing technique has yet been used to analyze its pharmacological properties. The drug-likeness of RRx-001 is proven by its physicochemical and ADMET properties, where it is projected to be passively absorbed by the gastrointestinal system and permeate the CNS without the use of P-glycoproteins. The Gene Ontology, KEGG pathways, and gene functions of the determined targets revealed that they are involved in cellular processes related to cancer pathways and radioresistance development. Two genes (CAMK2A and HSP90AA1) are linked in the cellular response to heat, while four of the query genes (VEGFA, CAMK2A, GPD1L, and HSP90AA1) contribute to peptidyl-serine modification. Ultimately, four candidates (VEGFA, HSP90AA1, CCND1, and ZEB2) showed the most significant protein-protein interactions among the hundreds of drug targets screened. These were verified by the inverse molecular docking result, which shows good binding affinities. This study provides the first comprehensive *in silico* target fishing evidence to show that RRx-001 is a promising drug candidate and radioprotective agent, targeting multiple proteins and pathways involved in cancer progression and radioresistance.

IN SILICO DESIGN OF QUERCETIN DERIVATIVES WITH POTENTIAL DUAL INHIBITORY ACTIVITY AGAINST GSK3*β* AND CDK5/P25 FOR THE TREATMENT OF ALZHEIMER'S DISEASE.

***A. Latorre*[1], *G. Nájera*[1], *J. Tamayo*[1], *J. Zavaleta*[2]**

[1]*Universidad Nacional Mayor de San Marcos, Peru*
[2]*Universidad Privada Antenor Orrego, Peru*

There are over 55 million people worldwide living with dementia in 2022, reports from the National Institute of Aging indicate that the prevalence of Alzheimer's disease (AD) doubles every five years. To assess existing strategies for compounds targeting AD, the current approach is the *in silico* design of multi-target drugs covering diverse pathways related to AD pathophysiology. Tau protein hyperphosphorylation mechanism associated with *β*-amyloid accumulation is a critical step in the pathogenesis and progression of AD, related to kinases GSK3*β* and CDK5/p25 activation. In this sense, the search for dual inhibitors is a novel therapeutic strategy for the treatment of AD[1]. The flavonoid scaffold has been the basis for the design of new derivatives as drug candidates against AD. Despite scarce studies with substitutions at the 8 Carbon position, they have shown to be great inhibitors of AD targets[2]. Herein, we described the *in silico* design of quercetin derivates substituted at the C8 position as dual inhibitors of GSK3*β* and CDK5/p25 through an integrative approach of physicochemical and toxicologic properties along with structural bioinformatics. Based on this integrative workflow, 613 bioisosteres were designed using the Swisbioisostere platform. The first screening was performed based on criteria of kinase inhibitory activity, toxicity, Lipinski-Veber rules, and pharmacokinetics, using Way2Drug, DataWarrior, and SwissADME platforms, respectively. 25 filtered bioisosteres were optimized with Avogadro and finally, site-specific docking was performed in the active site of GSK3*β* and CDK5/p25 (PDB ID: 1Q3D and 7VDP) using MTiAutoDock (https://acortar.link/1VTpZA). 10 compounds turned out to have better affinity energies to GSK3*β* and CDK5/p25 in comparison with quercetin (-9,8; -10.62 kcal/mol) and indirubin (-9.26; -10.43 kcal/mol) as controls, particularly compounds QT94 (-12.36; -13.06 kcal/mol) and QT112 (-12.23; -13.81 kcal/mol). Moreover, all of them showed many hydrophobic interactions with important protein residues from the active site of GSK3*β* (Ile62, Val70, Ala83, Leu188, Leu132, Asp133, Tyr134, Val135 and Pro136) and CDK5/p25 (Ile10, Phe80 and Lys133). In the same way, the bioisosteres had many polar interactions with both GSK3*β* (Val135 and Asp133) and CDK5/p25 (Cys83 and Asp86). Additionally, the interaction of the evaluated compounds with residues Leu 132, Tyr134, Arg141 and Cys199 could indicate a higher binding selectivity to GSK3*β* compared to its isoform GSK3*α*, since they are specific residues. In conclusion, C8-substituted quercetin derivatives proved to be potential dual inhibitors of GSK3*β* and CDK5/p25 for the treatment of AD.



QT94 interactions with GSK3*β* (A), QT94 interactions with CDK5/p25 (B)

1. Xie, H, et al (2017), *Oncotarget*, 8(11), 18118–18128.
2. Pitchai, Arjun et al (2020), *Heliyon*, 6(9).

CONSENSUS VIRTUAL SCREENING OF NATURAL PRODUCT DERIVATIVES AGAINST TUBULIN

***E. López-López*[1,2], *J.L. Medina-Franco*[2], *C.M. Cerda-García-Rojas*[1]**

[1]*Departamento de Química y Programa de Posgrado en Farmacología, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Mexico City, Mexico*
[2]*DIFACQUIM Research Group, Departamento de Farmacia, Facultad de Química, Universidad Nacional Autónoma de México, Mexico City, Mexico*

Modification of the dynamics of the tubulin-microtubule (Tub-Mts) system has generated effective strategies to treat different types of cancer. A large body of data, obtained from many studies of Tub-Mts inhibitors including clinical trials, supports that this system is a validated target [1,2]. There is a great structural diversity of compounds that interact with the Tub-Mts system, among the most common are the alkaloids derived from *Vinca* species, taxanes, derivatives of colchicine, and derivatives of laulimalide. Each one of these groups has its respective binding site [3]. A high proportion of Tub-Mts inhibitors has been developed from representative and common scaffolds that cover a definite region of the chemical space. Additionally, the majority of reported compounds have been associated with a limited ADMET profile. Hence, the complexity in rationally modeling or designing new agents that interact in the Tub-Mts system arises.

This study presents a virtual screening protocol for natural products with a consensus approach (ligand-based design + structure-based design) whose function is to identify potential inhibitors of the Tub-Mts system. A combined strategy of molecular similarity, docking, and molecular dynamics calculations, as well as ADMET property calculations, were used to identify new compounds to be prioritized and selected for subsequent biological evaluations [5,6]. The consensus method was constructed using the structural and chemical information of 851 compounds with described activity against the Tub-Mts system and reported cytotoxic activity in cancer cell lines.

Application of this approach allowed the selection of a group of compounds for future biological tests with a high probability of success as potential inhibitors of the Tub-Mts system from a database of 429 natural products isolated from plants that grow in the American continent. The selected compounds are derivatives of dihydroresveratrol, parvifoline, benzyl benzoate, and centaureidine.

We are thankful for the financial support from the NUATEI (Nuevas Alternativas para el Tratamiento de Enfermedades Infecciosas) program of IBT-UNAM to purchase an MOE academic license. We also thank the School of Chemistry, UNAM for funding this research with a PAIP (Programa de Apoyo a la Investigación y Posgrado) grant.

1. Arnst, K. E., Banerjee, S., Chen, H., Deng, S., Hwang, D.-J., Li, W., Miller, D. D. (2019) *Med. Res. Rev.*, 39, 1398–1426.
2. Chang, C. H., Yu, F. Y., Wu, T. S., Wang, L. T., Liu, B. H. (2010) *Toxicol. Sci.*, 119, 84–92.
3. Naaz, F., Haider, M. R., M. R., Shafi, S., Yar, M. S. (2019) *Eur. J. Med. Chem.*, 171, 310–331.
4. Guo, Q., Zhang, H., Deng, Y., Zhai, S., Jiang, Z., Zhu, D., Wang, L. (2020) *Eur. J. Med. Chem.*, 196, 112328.
5. Silva-García, E. M., Cerda-García-Rojas, C. M., del Río, R. E., Joseph-Nathan, P. (2019) *J. Nat. Prod.*, 82, 840–849.
6. López-López, E., Cerda-García-Rojas, C. M., Medina-Franco, J. L. (2021) *Molecules*, 26, 2483.

UNSUPERVISED MACHINE LEARNING-BASED SCREENING OF VOLATILE LEADS TO TARGET T790M MUTATED EGFR

A. Mukherjee*, *K.S. Mukunthan

Department of Biotechnology, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, India

Lung Cancer is one of the most prevalent and a leading cause of cancer mortality globally. Non-small cell lung cancer (NSCLC) accounts for 85 per cent of all lung cancer cases, with epidermal growth factor receptor (EGFR) kinase overexpression being a prevalent aetiology [1]. The adverse side effects and potency of the existing EGFR inhibitors are dwindling due to drug resistance driven by the T790M mutation. We focused on volatile compounds retrieved from different chemical libraries, possessing low molecular weight, high vapour pressure, and are lipophilic. They possess a diverse range of scaffolds and are deemed potential candidates for lead development. On the other hand, poor pharmacokinetics is frequently a stumbling block during drug development. However, virtual screening, a computer-based approach, aids in discovering novel hits from large chemical libraries utilising information about the protein target or existing bioactive ligands. We implemented a principle component analysis (PCA), an unsupervised machine learning algorithm, to reduce the high-dimensional chemical space of 1042 volatile compounds and recognise the data patterns [2]. A correlation between the actives/known and unknown compounds was deduced based on their broad structural and physicochemical features. The compounds were clustered based on structural similarity using the k-means algorithm. Further, the Silhouette method validated the cluster consistency. The potential leads were screened by determining their pharmacokinetics and toxicity profiling. Moreover, precluding drug-drug interactions, which often result in serious side effects, by determining the most favoured sites of Cytochrome P450 mediated metabolism. The identified leads' molecular interaction was studied with the T790M mutated EGFR using the Molecular docking technique with Osimertinib as a control.

The study demonstrated the vigour of the unsupervised machine learning method in screening potential leads for targeting secondary mutation in EGFR driving NSCLC. The reduced high-dimensional volatile chemical space opened an avenue to cluster the compounds based on similarity. The pharmacokinetics and toxicity profiling eliminated 95% of the unfavourable leads. While only four compounds satisfied all the criteria for being designated as potential leads, averting adverse side effects. Compound No. 946 and No. 347 possessed better binding energy with the T790M mutated EGFR than the control and other two leads. Moreover, the molecular interactions of the complex revealed hydrogen bond interactions and the preponderance of hydrophobic interactions, stabilising the lead at the target site and may aid in modulating the binding affinity and lead potency for inhibiting the overexpressed and mutated EGFR in NSCLC pathogenesis.

This study was supported by Dr. TMA Pai Scholarship of Manipal Academy of Higher Education, Manipal.

1. Ibrahim, M. T., Uzairu, A.S., Uba, G. A. Shallangwa (2021) *Futur. J. Pharm. Sci*., 6, 55.
2. Awale, M., Reymond, L. (2016) *J. Cheminform.*, 8(1), 1-8. May 2016.

SOIL ECOTOXICITY PREDICTION AGAINST *FOLSOMIA CANDIDA* USING 2D-QSAR APPROACH

R. Paul*, *M. Chatterjee*, *K.Roy

Drug Theoretics and Cheminformatics Laboratory, Department of Pharmaceutical Technology, Jadavpur University, Kolkata, India

Soil ecotoxicological test is an essential tool for risk assessment of various xenobiotic chemicals. Such tests can be performed using soil invertebrates by exposing them to specified soil contaminating chemicals. Soil invertebrates provide various ecosystem services (i.e., soil transformations are beneficial for mankind). For example, soil invertebrates may influence mineralization of nutrients in soil organic matter (SOM), affecting the "soil fertility" – an important factor from the agricultural aspect. Hence, soil invertebrates serve as an outstanding biological indicator of terrestrial ecosystem and overall soil quality, considering their high sensitivity when compared to other indicators of soil quality (physical/chemical). Therefore, laboratory tests using invertebrates can be considered as the mainstay of ecotoxicological impact assessment. Quantitative and/ or qualitative results elicited from such tests help several regulatory authorities across the globe to determine the ecological risk level of substances and safe exposure limits for human and soil biota. Thus, such valuable information enables governmental regulatory authorities to control manufacturing output and sale of pesticides, to decide threshold limit for safe application of residues to agricultural soils, etc. However, laboratory tests (both *in vivo* and *in vitro*) are costly and time-consuming affairs and involve extensive use of animals. Hence, such tests cannot be extended entirely for predicting toxicity of novel compounds. As a result, an alternative *in-silico approach* of quantitative structure–activity relationships (QSARs) is used for environmental risk assessment for novel compounds, free of the exhaustive use of test animals under the REACH regulations in the EU. In this background, necessary limited data available from laboratory tests on the soil invertebrate *Folsomia candida* (C. name: Springtail) were collated from the database of ECOTOX (cfpub.epa.gov/ecotox). Data is collected for the endpoint - pEC50 only. Samples of total 45 chemical compounds were selected for which chemical descriptors were calculated for each compound. Then the whole dataset is split into a test dataset (11 compounds) and a training dataset (34 compounds), based on Euclidean Distance based approach. Using genetic algorithm, significant descriptors out of all descriptors pool were selected. Using these selected set of descriptors both in test set and training set, the Best Subset Selection software (http://teqip.jdvu.ac.in/QSAR_Tools/) was run which gave best possible combinations of limited number of descriptors based on desired linear model equation length from which four best models were selected based on their internal and external validation metrics. Four partial least squares (PLS) models were built based on those four multiple linear regressions (MLR) models. These four models were then used in Intelligent Consensus Predictor version 1.2 (PLS version) to get the final consensus model, built using best selection of predictions (compound-wise) from four 'qualified' individual models. Both internal and external validations metrics of this consensus model are well- balanced and within the acceptable range as per the OECD criteria which is reconfirmed by predictions made through the Chemical Read-Across method. From the aforementioned parameters, certain conclusion on general contribution can be made: The consensus model was found to be better than any previous model developed using data in this order. The read-across predictions have shown even better metric values than consensus prediction. It has been found out that presence of phosphate group, electron donor groups, molecular weight and polyhalogen substitution have significant impact on soil ecotoxicity. The soil ecotoxicological risk assessment of organic chemicals can therefore be prioritized by these markers. As a result, these models developed for a wide range of organic compounds can be applied to any new query compound for improved accuracy in prediction.

THE CONSENSUS ENSEMBLE NEURAL NETWORK MULTITARGET MODEL OF ANXIOLYTIC ACTIVITY

***M.A. Perfilev*[1,2], *P.M. Vassiliev*[2,1], *A.A. Spasov*[1], *A.N. Kochetkov*[2]**

[1]*Volgograd State Medical University, Department of Pharmacology and Bioinformatics, Volgograd, Russia*

[2]*Research Center for Innovative Medicines, Laboratory for Information Technologies in Pharmacology and Computer Modeling of Drugs, Volgograd, Russia*

Expanding the variety of anti-phobic, anxiolytic and antidepressant drugs may be difficult due to the general complexity of behavioral responses they induce, challenges in modelling animal responses, etc. Virtual screening of new chemical compounds using artificial neural networks will make it possible to find successfully promising potent compounds. Previous studies [1] described the methodology for constructing the consensus ensemble neural network multitarget model of RAGE inhibitory activity. The aim of this study is to provide evidence concerning the universality of the approach taken as exemplified by anxiolytic activity.

We studied 14 target CNS-expressed proteins playing an important role in anxiety disorders. Molecular mechanical (Marvin 20.19 software [2]) and quantum chemistry (MOPAC2016 software [3]) approaches were used to optimize 3D models of ligand structures. Ensemble docking into binding sites was performed using the AutoDock Vina 1.1.2 [4]. The construction of models and the selection of learning parameters for predicting anxiolytic activity was performed using the Statistica Neural Networks module of the Statistica 8.0 software [5] for multilayer feedforward perceptron neural networks. The training sample included 658 known compounds from the ChEMBL database with clustered values of anxiolytic activity. In the current study, we propose an automated iterative training scheme involving several cycles and an additional iteration with custom settings.

More than 136000 neural networks were trained. We constructed a consensus ensemble neural network multitarget classification model of anxiolytic activity of chemical compounds which involves 3 ensembles consisting of 7 neural networks in each ensemble, for *high*, *high or moderate*, *active*-potency levels. A total of 21 neural networks were constructed.

The final recognition accuracy of the resulting model was 92.6%, while potency prediction was 89.5%. For proper model validation we used 4 reference drugs. The potency of all reference drugs was calculated correctly.

The current research has shown that the iterative structure-dependent model of chemical compounds with anxiolytic activity is effective and can be successfully used in the directed search for new compounds with potent anxiolytic effects.

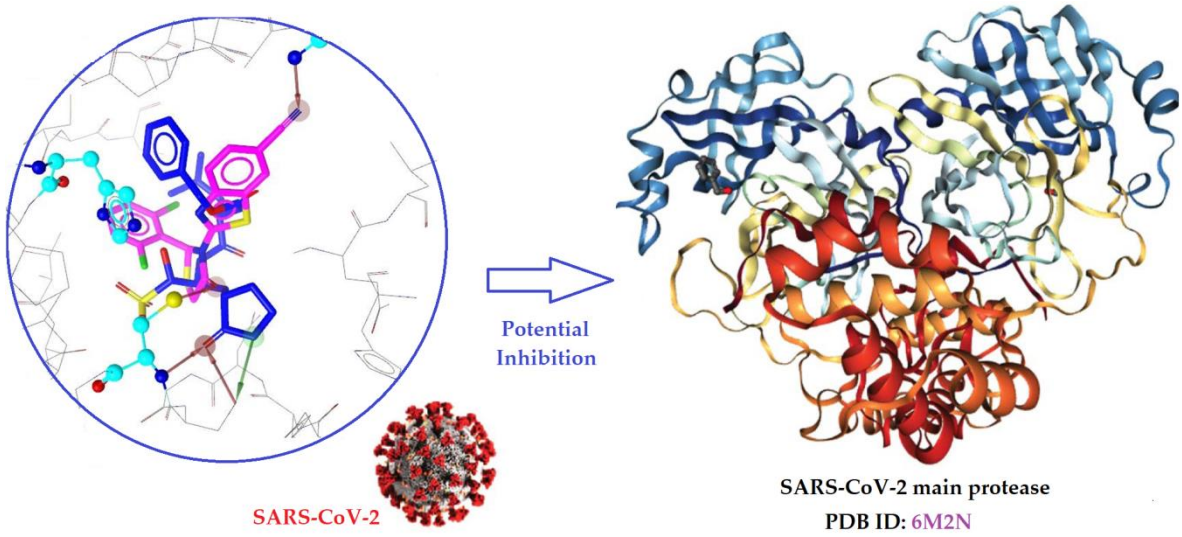
1. Vassiliev P.M., Spasov A.A. et al. (2021) *Biochemistry (Moscow)*, 67, 268-277.
2. ChemAxon: Marvin (2017) URL: https://chemaxon.com/products/marvin
3. MOPAC (2016) URL: http://openmopac.net/
4. Trott O., Olson A.J. (2010) *J. Comput. Chem.*, 31, 455-461
5. Statistica (2007), URL: http://www.statsoft.com/Products/STATISTICA-Features

THIAZOLE/THIADIAZOLE/BENZOTHIAZOLE BASED THIAZOLIDIN-4-ONE DERIVATIVES AS POTENTIAL INHIBITORS OF MAIN PROTEASE OF SARS-COV-2

***A. Petrou*[1], *P. Zagaliotis*[2,3], *N.F. Theodoroula*[3], *G.A. Mystridis*[3], *I.S. Vizirianakis*[3,4], *T.J. Walsh*[5], *A. Geronikaki*[1]**

[1]*School of Pharmacy, Faculty of Health Sciences, Aristotle University of Thessaloniki, Thessaloniki, Greece*
[2]*Division of Infectious Diseases, Weill Cornell Medicine, New York, USA*
[3]*Laboratory of Pharmacology, School of Pharmacy, Aristotle University of Thessaloniki, Thessaloniki, Greece*
[4] *Department of Life and Health Sciences, University of Nicosia, Nicosia, Cyprus*
[5]*Institute for Innovative Therapeutics and Diagnostics, Richmond, USA*

Since the time of its appearance COVID-19 has spread worldwide, with over 71 million confirmed cases and over 1.6 million deaths reported by the World Health Organization (WHO). Adding together, Delta and Omicron variants have also made the situation more challenging. Herein we report the evaluation of several thiazole/thiadiazole/benzothiazole based thiazolidinone derivatives which were chosen among 112 designed derivatives by docking as potential molecules to inhibit the main protease of SARS-CoV-2. The contained experimental data revealed that among the fifteen compounds chosen, five compounds (**k3**, **c1**, **n2**, **A2**, **A1**) showed inhibitory activity with IC_{50} within the range of 0.01 -34.4 μM. By assessing the cellular effects of these molecules, we observed that they also had the capacity to affect the cellular viability of human normal MRC-5 cells, albeit with a degree of variation. In particularly, the most promising compound, **k3,** with the higher inhibitory capacity to SARS-COV2 protease (0.01 μM) affects *in vitro* cellular viability only by 57% at the concentration of 0.01 μM after 48 h in culture. Overall, these data provide evidence on the potential antiviral activity of these molecules to inhibit the main protease of SARS-Cov2, a fact that sheds light on the chemical structure of the thiazole/thiadiazole/benzothiazole based thiazolidin-4-one derivatives as potential candidates for COVID-19 therapeutics.



STUDY OF THE EARLY STAGE STRUCTURAL REARRANGEMENT OF H3N2 INFLUENZA VIRUS HEMAGGLUTININ SUBUNIT HA2 DURING FUSION WITH A HOST CELL

***D.D. Podshivalov*[1], *E.M. Kirilin*[1,2], *S.I. Konnov*[1,3], *V.K. Švedas*[1,3]**

[1]*Research Computing Center, Lomonosov Moscow State University, Moscow, Russia*
[2]*Belozersky Institute of Physicochemical Biology, Lomonosov Moscow State University, Moscow, Russia*
[3]*Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russia*

Influenza virus hemagglutinin (HA) plays a key role in the process of virus and the host cell membranes fusion. Hemagglutinin includes lectin (HA1) and stem (HA2) domains. The stem domain of the HA is highly conservative compared to lectin domain. The binding of inhibitors in the hemagglutinin stem domain has been studied and X-ray structures of HA complexes with tert-butyl hydroquinone, N-cyclohexyltaurine, and the commercial drug umifenovir have been reported recently. It is hypothesized that the binding of these compounds could inhibit the structural rearrangement of the HA2 domain in the acidic environment of the endosome during virus entry to host cell and block the release of the influenza genetic material into the cell cytoplasm. *In silico* study of the hemagglutinin stem domain rearrangement can provide a detailed understanding of this process.

The early stage of the H3N2 influenza virus hemagglutinin stem domain structural rearrangement was studied in neutral and acidic conditions. All computer simulations were carried out for the entire HA2 domain. The protonation states of the amino acid residues at neutral and acidic pH were calculated using the PROPKA3 program. The additional constant-pH molecular dynamics simulation has been performed to refine protonation states for Glu120, Glu128, Glu131 and Asp132. Analysis of the electrostatic potential projections given by APBStools for pH 7 and 5 onto the hemagglutinin surface revealed change of charge at acidic conditions compared to neutral pH: the large interface area between three HA monomers near C-terminus switches its charge to positive. This region is formed by three symmetric parts, consisting of residues from 106 to 171. As a result of such charge changing the fusion peptide located in this region at the positively charged N-terminus is released, which initiates the hemagglutinin rearrangement process. Molecular dynamics in tandem with metadynamics was used to simulate the hemagglutinin HA2 domain structural rearrangement. The initial rearrangement stage is characterized by conformational changes in the C-terminal region of HA as well as changes in the position of the HA monomers side parts (residue 1 to 75), which are subsequently added as α-helices to stable regions (residue 76 to 105). The radius of gyration (RG) for all atoms of 106 to 171 amino acid residues was used as the first collective variable (CV) for the metadynamics to describe the rearrangement of the C-terminus area. The increasing RG indicates the increasing distance between three symmetric parts of HA2, which numerically describes this part of the process without any other variables. The distance (D_{com}) between the centers of mass of the side α-helix from 37 to 56 residues and the common center of mass of α-helices from 97 to 115 residues of two adjacent monomers has been chosen as appropriate second CV of the process. The free energy surfaces of the HA2 domain initial stage rearrangement were obtained after metadynamics simulation with the selected CV (RG and D_{com}) at pH 7 and 5. A comparison of the surfaces at neutral and acidic pH showed that changes in the conformation of the HA stem domain occur arbitrarily only after a change in the charge at the interface between monomers at low pH.

The developed modeling pipeline for studying the early stage of influenza virus hemagglutinin stem domain rearrangement using metadynamics makes it possible to simulate the process under various pH conditions. The proposed simulation scheme will allow studying the early rearrangement of hemagglutinin in complex with potential inhibitors and evaluating their effectiveness in the rational design and/or testing candidate molecules based on virtual screening.

The research was carried out using the equipment of the shared research facilities of HPC computing resources at Lomonosov Moscow State University [1]

This study was supported by the RSF grant No. 21-71-30003.

1. Voevodin, V.V., Antonov, A.S. et al. (2019) *Supercomputing Frontiersand Innovations* 6(2), 4–11.

STRUCTURAL DETERMINANTS OF FAMILY I SUBFAMILIES OF SOLUBLE INORGANIC PYROPHOSPHATASES

***K. Roshchin*[1], *R. Romanov*[2]**

[1]*Russian State University named after A.N. Kosygin, Moscow, Russia*
[2]*Center for Magnetic Tomography and Spectroscopy Lomonosov Moscow State University, Department of Fundamental Medicine Moscow, Russia*

Soluble inorganic pyrophosphatases family I (PPases) are an important object for study in biochemistry and molecular biology in order to develop new medicinal compounds. [1] Despite the fact that the enzymes of this family are well studied, the literature contains poor data on their taxonomy and structural features. [2] The inorganic pyrophosphatase from *Mycobacterium tuberculosis* (Mt-PPase) is a potential target for the development of anti-tuberculosis compounds and a popular subject for basic research. [3].

In this research, we use bioinformatics tools to search structural determinants of family I subfamilies of soluble inorganic pyrophosphatases. Our goal: to identify structural features of Mt-PPase. We used a software package: Mustguseal platform and Zebra2[4].

We selected 32 three-dimensional (3D) structures of PPases from various organisms. An array of 5186 enzyme sequences was obtained from the databases by BLAST procedures according to the reference structures. The array was clustered by Zebra2 in unusual iterative methodology.

According to the results of clustering, Mt-PPase fell into a cluster that also contains Inorganic pyrophosphatase from *Thermus Thermophilus* (2PRD) and Inorganic pyrophosphatase from Mycobacterium leprae (4ECP). One of the other clusters includes a well-studied protein - inorganic pyrophosphatase from *Escherichia coli* (Ec-PPase), which differs greatly in properties from Mt-PPase. Analysis of three-dimensional (3d) structures of Mt-PPase and Ec-PPase showed that the family-specific positions found by Zebra2 may play an important role in the formation of the properties of these enzymes.

Obtained results can be used to develop selective inhibitors of Mt-PPase.

Previously undescribed structure-property relationships have been found in the family I inorganic pyrophosphatases.

1. Kajander, T., Kellosalo, J., Goldman, A. (2013) *FEBS letters*, 587(13), 1863-1869.
2. Samygina, V. R. (2016) *Russian Chemical Reviews*, 85(5), 464.
3. Vijayan, M. (2005) *Tuberculosis*, 85(5-6),. 357-66.
4. Suplatov, D. et al. (2020) *Nucleic acids research*, 48(W1), W65-W71.
5. Romanov, R.S. et al. (2020) *Biochemistry (Moscow)*, 85(3), 326-333.

SMMOLE: PIPELINE FOR SEARCHING BIOLOGICAL PROPERTIES OF SECONDARY METABOLITES BASED ON THEIR MOLECULAR STRUCTURES

A. Sadovskaya*, *A. Gurevich

Center for Algorithmic Biotechnology, St. Petersburg State University, St. Petersburg, Russia

Secondary metabolites (SMs) are a rich natural source of industrially and medically important compounds. Modern mass spectrometry and genome sequencing technologies enabled collection of vast amounts of SM data from various environments. Researchers rely on computational methods to find novel molecules in these data [1-3]. However, such methods predict numerous error-prone SM structures that require further experimental validation. Creating an automated strategy for selecting the most promising *in silico* predicted molecules remains a bottleneck towards the high-throughput SM discovery.

Here, we present SMMole – a computational pipeline for assigning predicted SM molecules with tentative biological properties, such as producing organism and bioactivity. Researchers can utilize SMMole output for (i) eliminating SMs associated with producers that are inconsistent with the known sample environment, and (ii) prioritizing molecules with the strongest bioactivity for experimental validation. Our tool overcame the limitations of the competing approaches: the PubChem chemical search interface [4] (lack of information for SMs absent in PubChem) and PASS Online [5] (no taxonomy predictions). Moreover, SMMole is suitable for the high-throughput analysis since it allows the batch processing of hundreds of SMs.

Our pipeline takes as input molecules in standard chemical formats (SMILES, MDL MOL and SDF), converts them to SMILES, and queries PubChem for exactly these and all structurally similar compounds. The users control the similarity level by setting SMMole's threshold on the Tanimoto coefficient. Since we focus on natural products only, the pipeline computes the NP-likeness score [6] for all PubChem hits and filters outs likely synthetic compounds. SMMole obtains taxonomy data and biological tests results for the remaining molecules, and summarizes them for each input SM. The tool output contains the most common producer taxonomy rank and averaged data on bioactivity.

SMMole v0.1 is available as a command-line tool. We tested the pipeline on 550 compounds from MIBiG, a manually curated database of SMs, and it correctly identified taxonomy and bioactivity of 84% of the molecules (460). We plan to integrate SMMole with the leading high-throughput SM discovery methods [1-3] to complement their output with the tentative biological properties of the identified compounds. We anticipate our pipeline will facilitate the search for new therapeutic agents by helping researchers to select and prioritize *in silico* predicted compounds for experimental validation.

This study was supported by the Russian Science Foundation (grant No. 20-74-00032).

1. Behsaz, B. et al. (2020) *Cell systems*, 10, 99-108.
2. Cao, L. et al. (2019) *Cell systems*, 9, 600-608.
3. Behsaz, B. et al. (2021) *Nature Communications*, 12, 3225.
4. Kim, S. et al. (2020) *Nucleic Acids Research*, 49, D1388-D1395.
5. Filimonov, D. A. et al. (2014) *Chemistry of Heterocyclic Compounds*, 50, 444–457.
6. Ertl, P. et al. (2007) *Journal of Chemical Information and Modeling*, 48, 68–74.

MULTI-TARGET APPROACH ON *LEISHMANIA DONOVANI* AND FINDING OUT POTENT INHIBITORS FOR ESSENTIAL ENZYMES

D. Saha*, *A.N. Jha

Tezpur University, Tezpur, India

Multi-target approach is considered as notable in suppressing pathogen. Multi-target drugs developed by this approach selects multiple targets leading to reduced resistance, better safety and efficacy. Visceral Leishmaniasis (VL) also known as kala-azar is the most severe form caused by the sp. *Leishmania donovani*. VL has impacted Asian countries like Nepal, Bangladesh and some states of India such as Bihar, West Bengal and Uttar Pradesh. Selecting essential dual protein targets, adenine phosphoribosyl-transferase (APRT) and dihydroorotate dehydrogenase (DHODH) of purine and pyrimidine pathways of *L. donovani* respectively, have freshly come up as a multi-target approach towards Visceral Leishmaniasis. Adenine phosphoribosyltransferase (APRT) an enzyme used by *L. donovani* in scavenging purine from the host. APRT present in parasite's infective amastigote form plays an essential role in purine salvage. K_m value of *L. donovani* APRT for adenine and AMP is smaller than that of the Km value of human APRT enzyme. Contrast in K_m value of APRT may considered as a potential target for treating visceral leishmaniasis. Pyrimidine synthesis by *de novo* pyrimidine pathway is essential for cell growth and cellular metabolism specially DNA and RNA biosynthesis. In the pathway, DHODH participates in the fourth step which catalyzes to convert stereoselective oxidation of (S)-dihydroorotate to orotate. DHODH enzyme is found in various organisms including humans which divided into class 1 and 2 in which class 1 further subdivided into class 1A and 1B. In class 1A DHODHs, orotate release determine the overall rate limiting process whereas in class 2 DHODHs, orotate release is slow and also not capable in catalysis. This is because of the difference in terms of flexibility of catalytic loops among class 1A and class 2 DHODHs.

When we compared the APRT in *L. donovani* and human, it was found that *L. donovani* APRT contains 236 amino acids whereas human APRT has only 180 residues. Further, the location of APRT gene was also different as in *L. donovani* it is at 26^{th} chromosome and in human it is at 16^{th} chromosome. The sequence identity and similarity of *L. donovani* APRT and human APRT protein was 27% and 42.6%, respectively. The amino acid residues in *L. donovani* and human DHODH is 314 and 367, respectively. Sequence identity of *L. donovani* DHODH and human DHODH protein was found to be 23% whereas similarity was 37%. The total number of amino acid residues of *L. donovani* DHODH is 314 whereas the total residues of human DHODH is 367. The gene of *L. donovani* DHODH and human DHODH is present at chromosome 16 in both. Established inhibitors on the selected proteins with derivative compounds, virtual screening and molecular docking were done to find promising candidates for the disease inhibition. The investigation was initiated by collecting a total of non-redundant 6,220 compounds from PubChem database. The compounds were filtered through Lipinski rule and Veber's rule which later went for ADMET screening having 10 parameters. A number of 203 compounds which cleared all the 10 parameters were passed. In the final stage, the biological activities of compounds were predicted using PASS analysis, having antileishmanial activity. To know the binding interaction, molecular docking between APRT and DHODH along with 16 screened compounds were done. Docking score for APRT and DHODH with Ligand 2 and Ligand 3 showed better result on three different docking software. For detailed binding mechanism, molecular dynamics (MD) simulation was done to know the dynamic behavior of the protein in the presence of ligands. MD simulation of 20 ns was carried out for APRT and DHODH with top five ligands to visualize the stability of protein through-out the simulation. Further, two ligands which showed better result from the five ligands analysis will go for 100 ns MD simulation. Thus, multitarget drugs may have higher potential than single-target drug in Visceral Leishmaniasis therapy regimen.

COMPUTATIONAL DESIGN OF ANTHRACYCLINE-TYPE BIOISOSTERES WITH POTENTIAL ANTI-CANCER ACTIVITY

***M. Santander*[1], *J. Alvarado*[1], *L. Córdova*[2], *A Reyes*[3]**

[1]*Facultad de Estudios Superiores Zaragoza, UNAM, México*
[2]*Facultad de Medicina, UNAM, México*
[3]*Centro de Física Aplicada y Tecnología Avanzada, UNAM, México*

Cancer is one of the main high-mortality conditions worldwide, the death toll as of 2020 was estimated at ten million. Additionally, cancer has a significant economic impact, with the annual economic cost of the disease estimated at US$1.16 trillion in 2010.

There is a wide variety of anticancer drugs, such as alkylating agents, antimetabolites, microtubule stabilizing and destabilizing agents, hormonal antagonists, protein kinase inhibitors and anticancer antibiotic agents, within this last group we have the anthracyclines, on which our research project is focused.

A large body of published data demonstrates that the relevant pharmacological target of anthracyclines for their therapeutic activity is topoisomerase II. Anthracyclines have drawbacks associated with their toxicity, since they present cardiotoxicity, which is the main limitation in their pharmacological use, in addition, they are known to be irritants and induce mucositis.

Based on the above, the aim of this project was to find new molecules with anticancer activity using computational techniques. Our study started from an anthracycline nucleus, the limitations were attenuated through an *in silico* study of a set of bioisosteres; since in bioisosterism, functional groups of the pharmacological molecule are replaced while maintaining similar biological properties and it is useful for improving the physicochemical, pharmacological, pharmacodynamic, pharmacokinetic or toxicological properties of drugs. In this way 200 molecules were generated, then through virtual screening and a well-founded modification of the structural basis seven molecules were found. Finally, these molecules were subjected to a molecular dynamics study.

Through what was previously described, compounds with minimal or null toxicological properties were found. Likewise, the binding affinities to topoisomerase II obtained in molecular dynamics were 6,8,11-trihydroxy-8-[2-(hydroxyamino)acetyl]-5,7,8,9,10,12-hexahydrotetracen-5-one (-374.49 kJ/mol), 2,5,12-trihydroxy-6,11-dioxo-N-(pyrimidin-4-yl)-1,2,3,4,6,11-hexahydrotetracene-2-carboxamide (-457.96 kJ/mol), 8-[3-(dimethylamino)propoxy]-6,8,11-trihydroxy-5,7,8,9,10,12-hexahydrotetracene-5,12-dione (105.51 kJ/mol), 8-[3-(dimethylamino)propoxy]-6,8,11-trihydroxy-5,7,8,9,10,12-hexahydrotetracen-5-one (-104.13 kJ/mol), 6,8,11-trihydroxy-8-[(methoxyimino)methyl]-5,7,8,9,10,12-hexahydrotetracen-5-one (-346.47 kJ/mol), 8-(2,3-dihydroxypropoxy)-6,8,11-trihydroxy-5,7,8,9,10,12-hexahydrotetracene-5,12-dione (-337.12 kJ/mol) and 8-(2,3-dihydroxypropoxy)-6,8,11-trihydroxy-5,7,8,9,10,12-hexahydrotetracen-5-one (-315.52 kJ/mol).

DRUG REPURPOSING DRIVEN NONMATERIAL FOR *PSEUDOMONAS AERUGINOSA* INFECTION

S. Sinha*, *J. Anand*, *K. Pant

Department of Biotechnology, Graphic era (deemed to be University), Clement Town, India

Pseudomonas aeruginosa is one of the most prevalent multidrug resistant pathogens and gruesome microbial species responsible for burn wound infection. With increasing emergence of pathogenicity, the phytocompounds of herbal plants have emerged as one of most prominent inhibitory agents against wound pathogens. Furthermore, drug repurposing of existing drugs/antibiotics is an alternative strategy to overcome antimicrobial resistance. In the present study, the computational approach is employed to screen out the most significant molecular interaction between efficient plant-based inhibitors, drugs/antibiotics with outer membrane proteins in *P. aeruginosa*. Further, owing to the wound healing efficacy of the nanomaterial-based dressings, the most significant combination of inhibitors was used for synthesis of nano-material. The nano-material was characterized using UV-Vis Spectrophotometer, zeta- potential, FT-IR, TEM, and validated for antimicrobial effect against *P. aeruginosa in vitro* via agar well diffusion assay and MIC.

The study identified alternative and potent drug combinations as nano-formulation which could serve as novel therapeutics against the burn wound infection.

UNVEILING THE POTENTIAL DRUG LIGANDS AGAINST VIRULENCE-RELATED HYPOTHETICAL PROTEIN IN *CRYPTOCOCCUS NEOFORMANS*: AN *IN SILICO* ANALYSIS APPROACH

R. Co[1], L. Lacandula[1], D.A. Tina[1], L.V. Alvarez[1,2]

[1]Department of Biology, College of Science, Polytechnic University of the Philippines, Manila, Philippines

[2]Research Management Office, Office of the Vice President for Research, Extension, Planning and Development, Polytechnic University of the Philippines, Manila, Philippines

The biological functions of hypothetical proteins are considered uncharacterized regardless of the prediction of their existence. This is supported by the rapid development of microbial sequence databases that enabled *in silico* genome comparisons to be used in finding novel targets early in the drug development process. However, although multiple researchers have looked into the hypothetical proteins derived from various microorganisms, the field of fungal zoonoses remains understudied. This study aimed to screen the hypothetical proteins of the globally distributed zoonotic fungus *Cryptococcus neoformans* and discover the potential drug ligands of the most promising hypothetical protein. Among the 11,476 non-redundant hypothetical proteins in the Uniprot database, one final hypothetical protein, Q5KPC4, passed the qualifications of a good target. Using Interpro, this hypothetical protein was found to belong to the crotonases superfamily, which is a diverse group of target proteins, and is linked to the virulence and pathogenicity of *C. neoformans*. Using highly reliable *in silico* tools, Q5KPC4 showed flexibility and significant interactions, making it a suitable target for drug development. Moreover, five promising ligands were identified through the drugbank server. The properties of the determined potential drug ligands based on the ADMET and pharmacokinetic analysis imply that all ligands mirror the qualities of great drug candidates. Ultimately, the PASS online and molecular docking results revealed that the ligands have antifungal and anthelmintic activities that could serve as an antagonist, inhibitor, or substrate to Q5KC4. Hence, this study provides the first in silico evidence that thoroughly characterized the properties of Q5KPC4, denoting its association as a drug target.

COMPUTATIONAL ANALYSIS OF CADMIUM CHLORIDE INDUCED MICRORNA ASSOCIATED TOXICITY IN HK-2 CELLS

S.E. Vankamamidi, A. Mukherjee, K.S. Mukunthan

Department of Biotechnology, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, India

The carcinogenicity of cadmium chloride in humans has been extensively documented, and it is the most hazardous of the cadmium compounds. It is a reactive metal that can impair and damage the brain, heart, liver, kidneys, placenta, testicles, and other organs in mammals. Coughing, difficulty breathing, skin redness, stomach pain, diarrhoea, nausea, vomiting, and other symptoms are some of the symptoms. On a cellular level, ROS and LPO levels increase, causing cell stress and death. This toxicity's specific mechanism has yet to be established.

We hypothesize that miRNAs cause cadmium chloride-induced toxicity in the nephron's proximal tubular region. MicroRNAs suppress the expression of mRNA targets through promoting mRNA degradation and translational repression. The microarray data of HK-2 cell lines treated with cadmium chloride was analysed. Upon analysis we found out that the hub gene MRPS10 (Mitochondrial Ribosomal Protein S10), which codes for mitochondrial ribosomal protein (28S subunit protein) in mammals, was shown to be downregulated. Mitoribosome, which is found in mitochondria and is responsible for translating mitochondrial mRNAs encoded in mtDNA, contains this 28S subunit protein.

We could link this to our research since cadmium chloride can increase mitochondrial membrane permeability and reduce membrane potential, resulting in mitochondrial swelling. Superoxide dismutase, glutathione peroxidase, ATPase, and other ROS quenchers or scavengers were shown to be depleted, resulting in increased levels of reactive oxygen species (ROS) and lipid peroxidation (LPO). Overall, the mitochondrial pathway causes Cd-induced apoptosis [1]. There was a considerable overexpression of 38 miRNAs when proximal tubular cells were treated with cadmium chloride [2]. Using TargetScan and miRTarBase, we discovered that three of the 38 miRNAs directly targeted the gene MRPS10. We also believe that these three miRNAs are involved in the MRPS10 gene's downregulation.

The structure of the miRNA-mRNA duplex was found to be stable and its docking with the AGO protein (part of RISC silencing domain) was successful, and a good docking score was obtained. As this docking is successful, we can conclude that miRNAs play a role in the cadmium chloride toxicity pathway.

1. Mao, W P et al. (2011) *Human & experimental toxicology*, 30(8), 920-929.
2. Lemaire, J., et al. (2020) *International Journal of Toxicology*, 39(2), 103–114.

REPURPOSING OF FDA-DRUGS AS POTENTIAL ERβ AGONISTS USING MULTICOMPLEX-BASED PHARMACOPHORE MAPS. A NEW APPROACH IN BREAST CANCER THERAPY

***L.H. Vazquez-Mendoza*, J. Garduño-Durán, J. Correa-Basurto, H.L. Mendoza-Figueroa, J.B. García-Vazquez**

Laboratory for the Design and Development of Drugs and Biotechnological Innovation, Escuela Superior de Medicina del Instituto Politécnico Nacional , CDMX, México

Breast cancer is considered a global threat to women's health. The different therapies used for its treatment present disadvantages with respect to the efficacy and selectivity of anti-cancer drugs and the development of resistance to chemotherapy. The nuclear receptor ERβ has been of increasing interest as an important pharmacological target in cancer; the ERβ selective agonist DPN (diarypropionitrile) has been shown to inhibit the proliferation of breast cancer cells with decreased expression of Cyclin D1 and increased apoptosis and may also increase the expression of ERβ in cells with a down-regulation. Stressing further that clinically increased levels of ERβ result in a good prognosis and survival of patients with triple-negative breast cancer. Whereas the activation of this receptor represents an important therapeutic strategy in breast cancer, the goal of this work was the development and modeling of pharmacophoric maps based on multiple ligand-protein complexes using novel computational tools; a massive exploration was carried out in drug libraries such as DrugBank and FDA (used in the clinic for the treatment of another pathology) to identify those that meet the minimum characteristics necessary for selective modulation of the activity of ERβ, considered as an important receptor of recent pharmacological interest in the treatment of breast cancer. From this identification, virtual screening methods were used in combination with molecular docking simulations to optimize these results. Based on the results of the developed *in silico* strategy and data analysis, 10 drugs were obtained with the best evaluations. Of these drugs obtained, it was very useful to analyze their existing pharmacological information in relation to cancer, estrogen receptors, additional information such as therapeutic status, toxicological and physico-chemical properties that supported or disapproved of the choice as potential biological evaluation drugs. Finally, SB (211110-63-3) and LB (36894-69-6) were selected to be evaluated in vitro in breast cancer cell lines: MCF-7 and MDA-MB-231. An antiproliferative effect was observed with both drugs in the two cell lines; for SB a value of IC_{50} = 81.7 μM was obtained in MCF-7 and IC_{50} = 101.2 μM in MDA-MB-231, for LB drug a IC_{50} =127.7 μM was calculated in MCF-7 and IC_{50} =146.3 μM in MDA-MB-231 at 48h.

The development of a multicomplex based pharmacophoric map aimed at ERβ, using novel computational tools, allowed identifying by virtual screening and molecular docking, drugs with high affinity for the ERβ receptor, which demonstrated antiproliferative activity in the MCF-7 and MDA-MB-231 breast cancer cells. This poses a viable alternative for the possible repositioning of SB and LB, to be used in the therapy against luminal breast cancer and aggressive triple negative.

ABSTRACTS

MOLECULAR DOCKING OF URACIL DERIVATIVES INTO THE ACTIVE CENTER OF CYCLOOXYGENASE ISOFORMS (COX-1/COX-2)

***L.G. Abdrakhmanova*[1], *V.R. Khairullina*[2], *Y.Z. Martynova*[1], *A.G. Mustafin*[1,2]**

[1]*Department of Chemistry, Bashkir State University, Ufa, Russia*
[2]*Ufa Institute of Chemistry of the Russian Academy of Sciences, Ufa, Russia*

The steric complementarity of 24 pyrazole derivatives with active centers of cyclooxygenase–1 (COX-1) and cyclooxygenase–2 (COX-2) was studied by molecular docking using the AutoDock 4.2.6 program [1-2]. As models of COX-1 and COX-2 protein molecules, macromolecules with codes 3n8x (chain B) and 1PXX (chain A), respectively, were selected from the PDB. The active center of macromolecules was placed in a three-dimensional box with a size of 22 Å. The position of the native ligands — nimesulide and diclofenac for COX-1 and COX-2, respectively, was taken as the center of the box. The results are shown in Fig. 1.

a) b)

Figure 1. Results of compound positioning in the active site of COX-1 (a) and COX-2 (b)

The data indicate that compounds S1-S24 are comparable to the active components of known nonsteroidal anti-inflammatory drugs: nimesulide, diclofenac and flurbiprofen in terms of binding efficiency to the active centers of COX isoforms represented by the E_{bind} parameter within an error (2.5 kcal/mol). In particular, the numerical E_{bind} values for these compounds when studying their affinity with COX-1 and COX-2 active centers are in the range of -8.42 — -2.01 kcal/mol and -9.73 — -4.52 kcal/mol, respectively. A comparative analysis of numerical E_{bind} values for the same compounds calculated when studying affinity with COX-1 and COX-2 active centers suggests that S1-S24 compounds can detect a pronounced inhibitory effect on both COX isoforms in vivo, that is, they theoretically can be non-selective inhibitors of COX isoforms. The replacement of the hydrogen atom at the R1 position with alkyl fragments contributes to a slight increase in the numerical value of E_{bind} with respect to both isoforms of COX. An increase in the length of alkyl substituents at positions R1 and R2 also contributes to an increase in the numerical value of E_{bind} for both enzymes. This effect is especially pronounced for COX-2 inhibitors.

The reported study was funded by the Russian Science Foundation for Basic Research according to the research project No. 19-73-20073.

1. URL: https://www.rcsb.org/
2. URL: http://vina.scripps.edu/

INTEGRATIVE NETWORK ANALYSIS TO IDENTIFY POTENTIAL TARGETS AND ANTINEOPLASTIC LEADS FOR PANCREATIC CANCER

S. Abraham*, *A. Mukherjee*, *K.S. Mukunthan

Department of Biotechnology, Manipal Institute of Technology, Manipal Academy of Higher Education, Karnataka , India

Background. This study involved using a network pharmacology approach to identify promising natural therapeutic compounds to treat pancreatic cancer. Since 2000, the incidence of pancreatic cancer has increased by about 1% each year, accounting for 7% of all cancer fatalities [1]. In 2022, 62,210 new cases are estimated, accounting for 3.2 % of all new cancer cases [2]. Current studies on drug development for pancreatic cancer involve combined therapy comprising asparagine restriction and MAPK signalling inhibition [3]. Treatments approved for pancreatic cancer treatment include Gemcitabine, Fluorouracil injection, Abraxane, Olaparib etc. The goal of our research is to find novel natural alternatives to traditional medications.

Methods. FlavorDB, a database of flavor molecules, provided 456 potential drug molecules. Opuntia was selected as the source for lead identification. Each lead was optimized based on its ADMET properties using Maestro. The targets were obtained from GeneCards and BioJupies. Functional enrichment analysis was performed utilizing Metascape for the former and Bingo for the latter. The PPI network of the targets visualized using the string database was incorporated into Cytoscape to obtain the hub genes for docking purposes. Molecular docking was performed using GLIDE, a ligand docking tool in Maestro.

Results. Based on the findings of our research, lead 7 and lead 8 showed suitable binding affinity to the target proteins, 1D5R(PTEN) and RPS3 respectively. A docking score of -5.98 Kcal/mol was obtained when lead 7 was docked with 1D5R and a score of -3.8 Kcal/mol was obtained when lead 8 was docked with RPS3.

Discussion & Conclusion. PTEN is a common tumor suppressor gene and it's loss can be correlated with an increased vulnerability to this form of cancer [4]. Mutations are common for the PTEN gene. Any variation in this gene could induce changes in the structure of the protein it codes for, leading it to cease working perfectly or becoming inactive, resulting in aberrant cell proliferation. Lead 7 proved to be a good potential drug molecule capable of binding to PTEN and reversing this. Based on survival analysis using Kaplan-Meier Plot, it is understood that high expression of RPS3 decreases the survival of patients. Thus, lead 8 proved to be a suitable potential drug capable of binding to RPS3 to reduce expression.

1. URL: https://www.cancer.net/cancer-types/pancreaticcancer/statistics#:~:text=Incidence%20rates%20of%20pancreatic%20cancer,the%20United%20States%20this%20year
2. URL: https://seer.cancer.gov/statfacts/html/pancreas.html
3. Pathria G, Lee JS, Hasnis E, et al (2019) *Nature Cell Biology*, 21, 1590–1603.
4. Bazzichetto C, Conciatori F, Pallocca M, et al (2019) *Cancers*, 11, 435.

COMPUTATIONAL IDENTIFICATION OF ARSENIC TRIOXIDE INDUCED miRNA FOR ANTINEOPLATICITY

***P. Acharya*[1], *K.S. Mukunthan*[2]**

[1]*Department of Biotechnology, Patkar Varde College, Mumbai, India*
[2]*Department of Biotechnology, Manipal Institute of technology, Manipal academy of higher education, Manipal, India*

The incidence and mortality rates for carcinoma measure increasing by close to three-dimensional annually. Hepatocellular carcinoma (HCC) is that the major histologic form of carcinoma and accounts for roughly eightieth of the full carcinoma burden worldwide [1]. HCC is the most common liver malignancy and can be a main cause for cancer-associated dying worldwide [2]. As miRNA`s are involved in fine-tuning physiological responses and regulating gene expression by mediating post transcriptional silencing of target gene they have gained attention in diagnosis and therapy in number of disease [3]. The aim of the study was to identify miRNA related to hepatocellular carcinoma and identify their function and to use them as therapeutics in combination with arsenic tri-oxide. In silico approach was used for carrying out the experiment. Gene dataset of 371 genes were retrieved. Functional enrichment showed that these genes were involved in steroid biosynthesis process, cholesterol metabolic process, cellular response to cholesterol and related activities. **hsa-miR-4650-5p** showed to have cynical effect on the compounds responsible for cancer cell growth and tumour development. Secondary and tertiary structure of mRNA-miRNA duplex was studied. Molecular docking was carried out between AGO protein-mRNA-miRNA complex and their interaction was also studied. The present study identified the miRNA involved in HCC and its interaction with AGO protein and mRNA interaction to suppress cancer development at molecular level.

1. Nagy, Á., Lánczky, A., Menyhárt, O., Gyorffy, B. (2018) *Scientific Reports*, 8(1), 9227.
2. Balogh, J., Victor, D., Asham, E. H., Burroughs, S. G., Boktour, M., Saharia, A., Li, X., Ghobrial, M., Monsour, H. (2016) *Journal of Hepatocellular Carcinoma*, 3, 41–53.
3. URL: http://www.mirbase.org

IDENTIFICATION AND INVESTIGATING NOVEL CATECHIN BASED NANOCOMPOSITE HYDROGEL FOR INHIBITION OF WOUND INFECTION

***J. Anand*, N. Rai, S. Agarwal, A.Karnwal**

Graphic Era Deemed to be University, Clement Town, India

In case of wound healing, suitable wound dressings play an essential role in patient's care and hydrogel's have been identified as the ideal candidate which has been extensively employed to facilitate wound-healing processes, particularly burn wounds. However, further insight in developing and optimizing advanced hydrogel with antimicrobial properties offers a broad area of research in relation to specific clinical applications in wound healing.

Among the microbial contaminants of burn wounds, *Pseudomonas aeruginosa* is one of the most prevalent and gruesome microbial species responsible for burn wound infection and impairment of wound healing mechanism. Also, recent epidemiological studies have indicated antibiotic resistance in *P. aeruginosa* conferring it among the group of *ESKAPE* pathogens against whom alternative inhibitory agents are urgently required.

Green tea derived from *Camellia sinensis* plant is bestowed with multiple health benefits and its bioactive secondary metabolites are known to induce healing and regeneration of damaged tissues. Henceforth, considering the wound healing efficacy of hydrogel and antibacterial potency of major green tea, the current study highlights the efficacy of green tea encapsulate metal based nano-composite hydrogel, and investigating its inhibitory potential against siderophore biosynthesis in *Pseudomonas aeruginosa*.

Based on *in silico* investigation, the study identified potent alternative drug candidates targeting key emzymes of *P. aeruginosa* siderophore biosynthesis using iGEMDOCK. Further, copper nanoparticles encapsulated with selected potent candidate in synergistic concentration with catechins were synthesized and characterized by UV- Vis Spectruphotometry, EDX, FTIR, SEM, and validated for antimicrobial potential using MIC and Time kill assay. Further, nano-composite hydrogel was prepared and characterized for its thickness, transparency, swelling behavior. The present study unveils a novel therapeutic approach which can deliver a potential and safer nano-herbal synergistic antimicrobial formulation against other multidrug-resistant and pathogenic wound infection.

QSAR MODELING OF ACETYLCHOLINESTERASE INHIBITORS IN A SERIES OF METHOXYPYRIDINIUM DERIVATIVES

***K.G. Borkha Dorado*[1], *V.R. Khairullina*[2], *Y.Z. Martynova*[1], *A.R. Gimadieva*[2], *R.N. Nasretdinova, A.G. Mustafin*[1,2]**

[1]*Department of Chemistry, Bashkir State University, Ufa, Russia;*
[2]*Ufa Institute of Chemistry of the Russian Academy of Sciences, Ufa, Russia*

The purpose of this work was to establish quantitative relationship "structure-inhibitory activity" in a series of methoxypyridinium derivatives in relation to acetylcholinesterase (AChE) human, experimentally studied in the work [1], and building QSAR models based on these compound classes for virtual screening of virtual libraries and databases. The studies were carried out using the program GUSAR 2011 (General Unrestricted Structure Activity Relationships) according to the methodology, described in works [2]. As a result, six statistically significant stable consensus models were built QSAR. Each of the consensus models includes from 20 to 320 partial QSAR equations with statistical parameters R^2>0,7; Q^2>0,6 (Table 1). These models are applicable for virtual screening and search for new compounds with a pronounced inhibitory activity against (AChE) in order to develop new inhibitors of this enzyme. Additionally, structural descriptors have been identified that allow regulating the activity of inhibitors of this enzyme.

Table 1. Characteristics and prediction accuracy of IC_{50} values for consensus-models M1-M6

Training set	Models	N	R^2_{TrS}	R^2_{TS}	F	S.D.	Q^2	V
QSAR model based on QNA-descriptors								
TrS1	M1	301	0.834	0.7584	18.983	0.592	0.775	55
TrS2	M4	313	0.846	0.7364	20.328	0.566	0.792	59
QSAR model based on MNA-descriptors								
TrS1	M2	301	0.858	0.777	19.636	0.553	0.804	60
TrS2	M5	313	0.859	0.7195	20.874	0.545	0.810	61
QSAR model based on QNA- and MNA-descriptors								
TrS1	M3	301	0.859	0.7972	20.506	0.554	0.810	56
TrS2	M6	313	0.869	0.7564	22.688	0.532	0.822	58

N – number of structuresinthe training set; R^2_{TrS} - a multiple coefficient of determination calculated for compounds from the training set; R^2_{TS} - a multiple coefficient of determination calculated for compounds from the test set; Q^2 – a cross-validated R^2 calculated during leave-one-out cross-validation procedure on data of the training set; F – Fisher's coefficient; SD –standard deviation; V- the number of variablesin the finalregression equation

The reported study was funded by the Russian Science Foundation for Basic Research according to the research project No. 19-73-20073.

1. Lu, W., Xue, Y. (2010) *European Journal of Medicinal Chemistry*, 46, 1167-1172.
2. Khairullina, V., Safarova, I., Sharipova, G., Martynova, Y., Gerchikov, A. (2021) *Molecules*, 26, 421.

SYNTHESIS AND ACTIVITY OF 3,4-DIHYDROPYRIMIDIN-2-ONES(THIONS)

N. Ciobanu, F. Makaev

Institute of Chemistry, Laboratory of Organic Synthesis and Biopharmacy, Chisinau, Republic of Moldova

Chemistry is an experimental science, a science of practice, trial and error. All calculations must be confirmed by the results of studies of the structure of the substances obtained, must be verified by experience with subsequent recommendations for practical use. Modeling the properties and reactivity of chemical compounds is an integral part of the research strategy in general. This is especially true during the rapid development of science in general, with mutual cooperation and information complementarity. The main reasons are determined by the success of the development of theoretical ideas about the structure of substances, practical production and high achievements of computer technology. In chemistry, as in other areas of exact computational and practical sciences, the following areas can be distinguished: modeling the properties of molecules, dynamics and activity, point incorporation into the computational structure. Calculations of the structure and spectra of individual molecules are confirmed by modern methods of chemistry and theoretical molecular spectroscopy. At present, it is possible to make fairly reliable predictions and post hoc results for molecular systems and for each molecule separately. Modeling of chemical reactions of the first level, i.e. reactions of the first stage in a multistage process, or the so-called theoretical calculation of the structure and the ratio of the reactivity of the molecule to the reaction dynamics, does not always correspond to the expected result. Such chemical reactions, accompanied by a redistribution of their constituent parts, are still possible only for the simplest processes [1]. The process depends on the potential energy of the entire system determined by the given surface, in which the nuclei follow along classical trajectories.

In the synthesis of dihydropyrimidines, the goal is to select reagents and test various catalysts and conditions, especially in developing strategies to achieve the goals. Dihydropyrimidines have shown a wide range of pharmacological activity: analgesic, antibacterial, antihypertensive, etc., which makes further searches in their series very promising [2]. Monastrol (ethyl-6-methyl-4-(3-hydroxyphenyl)2-thioxo-1,2,3,4-tetrahydro-pyrimidine-5-carboxyl) is such a product of the dihydropyrimidines of the three-component reaction under Biginelli reaction conditions [3]. Therefore, the calculation of the flow conditions, condensation conditions, time, type and amount of catalyst, the presence of a solvent, what kind of solvent and its amount, it is possible to theoretically calculate the course of the reaction in one direction or another, and even the preliminary yield of the product. This is especially important, knowing that the synthesis under the conditions of the Biginelli reaction proceeds in more than five named directions, according to known schemes.

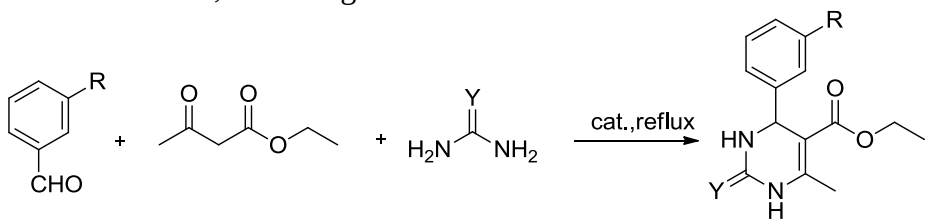


Fig.1. Scheme for obtaining dihydropyrimidines.
R: H(1), OH(2); Y: O(3), S(4);
5: Y=O, R=H; 6: Y=O, R=OH; 7: Y= S, R=H; 6: Y= S, R=OH

The work was supported by the applied research project 20.80009.5007.17 of the National Agency for Research and Development of Moldova.

1. Kapur, T.M., Maier, T.U., Kaflin, M.L. et al. (2000) *Journal of Cell Biology*, 5, 975-988.
2. Klein, E., DeBonis, S., Thiede, B. (2007) *Bioorganic & Medicinal Chemistry*, 19, 6474-648.
3. Kappe, O.C., Shishkin, O.V., Uraya, G., Verdinoa, P. (2017) *Neurotoxicity Research*, 07, 1859–1862.

RECONSTRUCTION OF GENE NETWORK ASSOCIATED WITH SCHIZOPHRENIA FOR GENE TARGETS SEARCH

***A.Y. Dokhoyan*[1], *M.V. Glushchenko*[1], *Y.L. Orlov*[1,2]**

[1]*First Moscow State Medical University, Moscow, Russia*
[2] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*

Schizophrenia is a severe polymorphic mental disorder (or group of mental disorders) characterized by a disruptions of thought processes and emotional reactions. Schizophrenia is a serious mental disorder, this is due to its steady growth in the population, its disabling effect on health, psychoemotional and professional life of patients; this disease is understudied, has a variety of symptoms typical for other pathological conditions, and the complexity of diagnosis without clear treatment. Using the on-line bioinformatics tools OMIM (omim.org), PANTHER (http://pantherdb.org/), DAVID, GeneMANIA, STRING-DB (https://string-db.org/) and GeneCards (https://www.genecards.org/) following the methods published in [1], we have analyzed the current array of data related to schizophrenia, calculated the categories of gene ontologies for a large list of genes, visualized them, and built gene networks containing the identified key objects and their relationships. Accumulation of genetic data in the field of schizophrenia research culminated in identifying risk factors and confident prediction of the disease occurrence. To find new gene-targets for diagnostics and therapy we have to reconstruct gene network of the disease, to cluster genes in the network, to reveal key (hub) genes with largest number of interactions in the network. Reconstruction and analysis of the structure of gene networks using bioinformatic methods makes it possible to identify key disease genes.

Using the OMIM resource, we obtained a list of genes associated with schizophrenia, which we used to search for gene ontologies using the PANTHER and DAVID databases. Based on the search results, we found that schizophrenia is associated with impaired synaptic signaling and intercellular signal transduction via neurotransmitters. With STRING-DB and GeneMANIA visualized and constructed gene networks, containing the identified key sites and their relationships, allowing to identify gene interactions; we found a strongly linked cluster that includes BDNF, SLC6A4, HTR2A, HTR2C, CHRM1, SRC, AKT, YWHAE, DISC1, DRD2, COMT, NDEL1, NOS1, CAMK28; with GeneCards the most relevant schizophrenia genes - COMT, DISC1, HTR2A, NRXN1 - were identified.

Reconstruction and analysis of the structure of gene networks with bioinformatic methods makes possible to identify key disease genes. However, translating the results into biological understanding is still a promising major challenge since schizophrenia is a genetically complex disease with a wide divergence of causes and conditions of onset.

1. Orlov, Y.L., Galieva, A.G., Orlova, N.G., Ivanova, E.N., Mozyleva, Y.A., Anashkina A.A. (2021). *Biomeditsinskaya Khimiya*, 67(3), 222-230.

ANALYSIS OF BIG MULTIDISCIPLINARY DATA FOR OPTIMIZATION OF THE DEVELOPMENT AND UTILIZATION OF PHARMACEUTICAL AGENTS AGAINST CORONAVIRUS INFECTION

D.S. Druzhilovskiy, P.I. Savosina, N.Yu. Biziukova, D.A. Filimonov, S.M. Ivanov, D.A. Karasev, P.V. Pogodin, L.A. Stolbov, O.A. Tarasova, A.V. Veselovsky, V.V. Poroikov

Institute of Biomedical Chemistry, Moscow, Russia

The COVID-19 pandemic is having a significant impact on all aspects of human life. During the past years, knowledge about SARS-CoV-2 coronavirus, pathological mechanisms, infectious process, and approaches to prevent and cure the disease has expanded significantly. However, the need for finding of new effective and safe anticoronavirus agents is still of current interest, as well as the need to improve approaches to the treatment of COVID-19 [1, 2].

We have developed the pilot version of the Informational-Computational System (ICS) as the AntiCOVID-19 portal freely available on the Internet (http://www.way2drug.com/anticovid/). Informational part of ICS provides information about the disease; data on biological processes in the human body affected by SARS-CoV-2 coronavirus; a list of molecular targets, which may be inhibited to block the development of the pathological process and/or reduce the risk of complications in COVID-19; a list of drugs recommended and being investigated as "candidates" for the treatment of COVID-19; links to web resources related to this subject area. Computational part of ICS includes modules for: (1) generating hypotheses regarding drug repositioning based on an assessment of structural similarity with respect to the pharmaceutical agent used as a query; (2) predicting anticoronavirus activity profiles based on classification models of structure-activity relationships, and (3) predicting biological activities associated with the prevention/reducing the risk of complications from COVID-19 (long COVID).

The possibilities of text mining approach based on various machine learning methods to select information about the names of pharmacological substances that are potentially effective in the treatment of a new coronavirus infection have been studied. We have proposed a new method for extracting chemical named entities (CNEs) from texts based on the naïve Bayes classifier [3]. In contrast to the earlier developed methods, our approach uses the representation of the data as a set of fragments of text (FoT) with the subsequent preparation of a set of multi-n-grams (sequences from one to n symbols) for each FoT. Since our approach does not require the usage of a predefined vocabulary, it may provide the recognition of novel CNEs. The average values of invariant accuracy range from 0.95 to 0.99 depending on the context window and the size values of n of the multi-n-grams. We applied the developed algorithm to the extracted CNEs of potential SARS-CoV-2 Mpro inhibitors and validated the obtained results. Manual analysis of the appropriate texts confirmed that CNEs of potential SARS-CoV-2 Mpro inhibitors were successfully identified by the proposed method.

Within the European initiative "JEDI Grand Challenge Against COVID-19" (https://www.jedi.foundation/covid19challenge) we performed the virtual screening of potential anti-SARS-CoV-2 agents among more than one billion molecules available for synthesis. Our team was included in the chosen 20 of the 160 research groups that took part in the project, which "hits" were selected for synthesis and testing of antiviral activity. As a result, 876 compounds have been synthesized and tested in *in vitro* assays (38 proposed by our team); and 28 active substances have been identified (one PLpro inhibitor proposed by our team). Our participation in this challenge provides the experience on retrieval of big chemical and biological data, to identify the potential anticoronavirus agents.

The study was supported in the framework of the Russian Foundation for Basic Research grant No. 20-04-60285.

1. Muratov, E.N., Amaro, R., Andrade, C.H. et al. (2021) *Chem. Soc. Rev.*, 50, 9121.
2. Kozlov, M. (2022) *Nature*, 601, 496.
3. Tarasova, O.A., Biziukova, N.Yu., Filimonov, D.A., Poroikov, V.V. (2022) *J. Cheminform.*, submitted.

SYNTHESIS AND IN SILICO ESTIMATION OF BIOLOGICAL PROPERTIES OF A NEW ALKYNE-CHOLESTEROL CONJUGATE

H. Pozniak[1], Y. Faletrov[1,2], V. Shkumatov[1,2]

[1]Research Institute for Physical Chemical Problems, Belarusian State University, Minsk, Belarus;
[2]Faculty of Chemistry, Belarusian State University, Minsk, Belarus

Alkyne derivatives of biomolecules have become popular tools due to their possibility to be detected inside cells using either CuAAC click chemistry [1] or Raman microscopy [2] approaches. Since cholesterol and hydroxycholesterols play many important roles as signaling molecules, labeled cholesterol analogues are used as molecular tools to understand better uptake, distribution and metabolism of the molecules [3, 4, 1]. A sterol derivative (CCF-EtAn) has been synthesized by conjugation of cholesteryl chloroformate with 3-ethynylaniline. CCF-EtAn purity and identity were confirmed using TLC (Rf=0,7, SiO_2, eluent benzene :ethanol 4:1) and ^{1}H-NMR (500 MHz, $CDCl_3$, δ = 6.59 (s, 1H), 5.46 (dd, J = 4.9, 2.4 Hz, 1H) *(B-ring =CH-)*, 4.65 (tt, J = 11.5, 4.8 Hz, 1H) *(alkyne)*.The structure has been considered to be new according to Pubchem database search, but structure of alkyne-cholesterol CID102257532 is its p-ethynyl analogue and CID102080258 is its superstructure. Biological properties of CCF-EtAn were evaluated *in silico* using PerMM server (for phospholipid membrane permeability) [5] and inverse high-throughput virtual screening using Autodock Vina [6] and a helper tool FYTdock [7]. 1000 randomly chosen PDB structures of sterol-binding proteins were used. We obtained the following PerMM calculation results (parameters: pH 7.35, T = 37 °C): Free energy of binding was equal to -10 kcal/mol, log of permeability coefficients for plasma membrane, BBB and Caco-2 models were equal to +3, -1.84, and -2.71, respectively. These results prove good permeability of phospholipid membranes for CCFEtAn. Docking results revealed many hits and top-5 includes PDB structures of cytochromes P450, namely, 5frb and 6q2t (CYP51) with energy of binding value (Ebind) -14.6 and -13.9, respectively, 3mzs and 3na0 (CYP11A1) with Ebind -14.3 and -14.0, respectively, as well as 6bd6 (CYP3A4) with Ebind-13,8.

This study was supported in the framework of GPSR (Belarus) No. 20210560.

1. Kel, A., Boyarskikh, U., Stegmaier, P. et al. (2019) *BMC Bioinformatics*, 20, 119.
1. Hofmann, K., Thiele, C., Schött H. et al. (2014) *J. Lipid Research*, 55, 583.
2. Yamaguchi, S., Matsushita, T., Izuta, S. et al. (2017) *Sci. Rep.*, 7, 41007.
3. Faletrov, Y.V., Efimova, V.S., Horetski, M.S. et al. (2020) *Chemistry and Physics of Lipids*, 227, 104850.
4. Faletrov, Y.V., Pozniak, H.I., Yakovets, P.S. et al. (2022) *Proceedings of the National Academy of Sciences of Belarus, Chemical series*, 58, 62.
5. Lomize, A.L., Pogozheva, I.D., Mosberg H.I. (2011) *J. Chem. Inf. Model.*, 51, 930.
6. Trott, O., Olson A.J. (2010) *J. Comp. Chem.*, 31, 455.
7.Faletrov, Y.V., Staravoitava, V.A., DudkoA.R. et al. (2022) PREPRINT (Version 1) available at Research Square [https://doi.org/10.21203/rs.3.rs-1456627/v1].

NEW POTENTIAL COVALENT INHIBITORS FOR SELECTED SARS-COV-2 PROTEINS

***Y. Faletrov*[1,2], *V. Staravoitava* [1,2], *V. Shkumatov*[1,2]**

[1]*Research Institute for Physical Chemical Problems, Belarusian State University, Minsk, Belarus*
[2]*Faculty of Chemistry, Belarusian State University, Minsk, Belarus*

The SARS-COV-2 pandemic is still a global threat and it might be danger for a long period of time if effective treatment will not be found. Thus, the discovery of new effective drugs is required to address the current COVID-19 health emergency and improve future chances to prevent coronaviral pandemics. High throughput virtual screening (HTVS) is often used to identify structures for further in vitro biological assays. Here we report about our in silico efforts to find unknown affine interacting pairs of few SARS-CoV-2 proteins and natural compounds/analogues with α,β-unsaturated carbonyls. Such compounds are known to be able to form covalent bonds with nucleophilic atoms of cysteine [1]. HTVS was done in inverse mode using a self-created helper software FYTdock, which employs the AutoDock Vina as well as known Binana.py script, oBabel executable files and original scripts and Microsoft Excel files with VB macroses [2]. Autodock Vina parameters were as follow: exhaustiveness 12, grid box cube side 4 nm. As a result, few protein-ligand complexes for two types of SARS-CoV-2 viral proteins with the lowest binding energies between -8 and -9.8 kcal/mol and also with an electrophilic atom within 0.4 nm of the cysteine sulfur atom were selected only. For NSP2 protein (pdb 7msw or 7msx) such pairs were found with physalin C (CID101650329) from Physalis, lobocrassolide (CID9975683) from coral species Lobophytum crissum, acetylivalin-like compound CID9884868 from Arnica longifolia, compound CID132585258 for which the formation of a covalent bond with cysteine residues of ubiquitin-conjugating enzyme is known [3], compounds CID15866740, CID154496991, CID21634938, CID5281446, CID51723030, CID11699341, CID15866746. For NSP10 (pdb 7diyc) the pair was found with lanceocrepidiaside F (CID21580547) from Crepidiastrum lanceolatum (Table 1).

Table 1. The protein-ligand complexes under consideration

Pdb code of protein	PubChem ID (CID), ligand name (if appropriate)	Ebind, kcal/mol	Targeted CYS residue
7msw	CID101650329, Physalin C	-9.8	CYS253
7msw	CID11699341	-9	CYS326
7msw	CID132585258, UbcH5c-IN-1	-8.4	CYS253
7msw	CID154496991	-8.3	CYS253
7msw	CID15866740	-9	CYS253
7msx	CID15866740	-8.5	CYS253
7msx	CID15866746	-8.5	CYS326
7diy	CID21580547, Lanceocrepidiaside F	-8.9	CYS74
7msw	CID21634938, Achalensolide	-8.3	CYS253
7msw	CID51723030, Berkedrimane B	-8.1	CYS253
7msx	CID5281446, Eupachlorin	-8.2	CYS253
7msw	CID9884868, Acetylivalin-like	-8.6	CYS253
7msw	CID9975683, Lobocrassolide	-8.8	CYS253

This study was supported in the framework of BRFFR X21COVID-030.

1. Serafim, R.A.M. et al. (2022) *J. Med. Chem.*, 65, 3173.
2. Faletrov, Y.V., Staravoitava, V.A., Dudko, A.R. et al. (2022) PREPRINT (Version 1) available at Research Square [https://doi.org/10.21203/rs.3.rs-1456627/v1].
3. Chen, H., Guozhen, W., Gao, S. et al. (2017) *J. Med. Chem.*, 60, 6828.

NITROBENZOXADIAZOL PNENYLBORONIC ACID: IN SILICO EVALUATION AS INHIBITOR FOR E. COLI ENZYMES AND IN VITRO TESTS WITH BOVINE ALBUMIN AND YEAST LIPASE

***Y. Faletrov*[1,2], *H. Pozniak*[1,2], *P. Yakovets*[1,2], *V. Shkumatov*[1,2]**

[1]*Research Institute for Physical Chemical Problems, Belarusian State University, Minsk, Belarus*
[2]*Faculty of Chemistry, Belarusian State University, Minsk, Belarus*

A hybrid compound joining 7-nitrobenzoxadiazol (NBD) fluorophore with diol-targeting phenylboronic acid group has been synthesized by reaction of commercially available NBD-chloride and 4-amino-3-fluorophenylboronic acid (AFPBA) in 1:1 mixture of ethanol and 50 mM aqueous sodium bicarbonate buffer at room temperature giving compound 1, NBDboronic1, as dark orange solid. The compound structure was in agreement with 1H-NMR (chemical shifts in D2O: 7,67-7,65 (dd, 1H), 7,29-7,24 (m, 2H), 7,09 (t, 1H), 6,92 (t, 1H) and 19F-NMR. Purity of 1 was confirmed by thing layer chromatography (SiO2, EtOAc: Rf 0.5 vs 0.8 for AFPBA). 1 was found to be a non-fluorescent compound with absorbance maximum at 460 nm in protic solvents (H2O, MeOH, EtOH) and 470 nm in aprotic (CHCl3, CH3CN). Absorbance maximum of 1 differs in aqueous 5 % NH_3 or CH_3COOH revealing the isosbestic point at 455 nm.

Development of new antibacterial drugs is an important task and boronic acid-based compounds are attractive as potential inhibitors for Ser/Thr-containing enzymes [1]. To estimate affinity and geometry of possible interaction of 1 with proteins of bacteria E. coli we performed inverse virtual screening for 8600 structures of proteins of the bacteria from Protein Data Bank (chains A were used only) using Autodock Vina (parameters: exhaustiveness 12, number of models 5) and an originally developed helper tool FYTdock to organize, run and analyze results of the screening [2]. The option allowed to complete the analysis for 16 days on a laptop (2.6 GHz Intel Core i5, 4 GB RAM). It was found that the docking score (Ebind) ranges from -11.7 for formate dehydrogenase (pdb 1kqg) to -2.7 for heat-labile enterotoxin (pdb 2xrs). Co-localization of boronic group of 1 with HO-group of Ser residues of the proteins were found in 635 cases and the lowest Ebind in the set was found to be -10.5 for acid phosphatase (pdb 1rmq). Analogously, co-localization of boronic group of 1 with HO-group of Thr residues of the proteins were found in 490 cases and the lowest Ebind in the set was found to be -10.4 for formate dehydrogenase (pdb 1fdi). Analysis of subset of 200 beta-lactamase structures allowed to find 60 poses with boronate-Ser co-localization with Ebind from -9.4 (pdb 4zj1) to -7.5. Analysis of subset of 35 penicillin-binding proteins structures revealed 6 poses with boronate-Ser co-localization with Ebind from -9.6 (pdb 5fgz) to -7.4. In vitro test with E. coli on a minimal salt medium allowed to find that 1 is able to reduce nitrobluetetrazolium reduction at 100 mkM. Also 1 was found to inhibit fluorescein dihexanoate hydrolysis by lipase from Candida rugosa (by 8 times at 100 mkM) and to be able to quench Trp fluorescence of bovine serum albumin with apparent dissociation constant ~ 20 mkM.

The work was supports by Belarusian State University grant № 20211802.

1. Song S. et al. (2021) *Acta. Pharm. Sin. B.*, 11, 3035.
2. Faletrov, Y.V., Staravoitava, V.A., Dudko A.R. et al. (2022) PREPRINT (Version 1) available at Research Square [https://doi.org/10.21203/rs.3.rs-1456627/v1].

DIFFERENTIAL PROFILING OF MOLECULAR NETWORKS AND PATHWAYS ASSOCIATED WITH THE OCCURRENCE OF MASTITIS IN DAIRY COWS

B. Ganguly

Research & Development Division, Ayurvet Limited, Baddi, India

Mastitis is the costliest veterinary disease, responsible for serious economic losses to dairy enterprises and posing major public health concerns. The disease causes direct losses in the form of morbidity and loss of yield as well as indirect losses through increased costs towards veterinary services and increased frequency of replacement of cows. Mastitis accounts for the greatest use of antibiotics among all veterinary diseases. Microorganisms that commonly cause bovine mastitis can also cause zoonotic diseases like scarlet fever, sore throat, diarrhea, *etc*. Shedding of toxins through milk by the mastitis-causing organisms is another reason for public health concerns. Cows exhibit significant differences in resistance as well as response to mastitis. The RNA-seq dataseries GSE93082 [1] of the milk cells from mastitic and non-mastitic, healthy cows were compared to identify prominent differences in terms of biomolecules involved, and networks and pathways thereof. GEO2R [2] was used to identify the differentially-expressed genes; the network was modeled using STRING and visualized in Cytoscape. STRING was also used to identify the salient pathways involved in the networks.

The performed analysis revealed that the milk cells of mastitic cows had rather different transcriptional profiles that involve fairly different although partially overlapping sets of master regulators in the molecular circuits that control activity of the gene expression. Based on the revealed master regulators, the results of this study may foster developments in marker-based genetic selection, diagnosis and prognosis; and target-oriented therapy of bovine mastitis.

The author was financially supported during this study by M/s Ayurvet Limited, India.

1. URL: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE93082; Accessed on 14.04.2022.
2. Barrett, T., Wilhite, S. E., Ledoux, P. et al. (2012) *Nucleic acids research*, 41, D991.
3. Szklarczyk, D., Morris, J. H., Cook, H. et al. (2016) *Nucleic acids research*, gkw937.
4. Smoot, M. E., Ono, K., Ruscheinski, J. et al. (2011) *Bioinformatics*, 27, 431-432.

P-GLYCOPROTEIN-ORIENTED OPTIMIZATION OF MDM2 INHIBITORS

***T. Grigoreva*, D. Novikova, V. Tribulovich**

Saint Petersburg State Institute of Technology, St. Petersburg, Russia

P-glycoprotein is of considerable interest for the design of drugs capable of treating chemoresistant tumors. This transporter possesses a low specificity; its substrates include various endogenous substrates, dyes, and drugs.

Some E3 ligase MDM2 inhibitors used to trigger apoptosis in tumor cells facilitate overcoming multidrug resistance by suppressing efflux by P-glycoprotein. In this work, we performed screening of own database of MDM2 inhibitors based on indole and isoindolinone in order to identify the most effective P-glycoprotein inhibitors among them.

The transporter consists of two pseudosymmetrical parts, and each includes the cytoplasmic ATP-binding domain and six hydrophobic transmembrane alpha-helices, that form the substrate translocation pathway. We considered the interaction of compounds with the transmembrane domain similar to known inhibitors such as tariquidar.

Three-dimensional structures of small molecules were generated using the LigPrep instrument in the OPLS3e force field with the generation of tautomers and stereoisomers. The structure of the RCSB Protein Data Bank PDB ID: 4Q9H was prepared using instruments of Protein PrepWizard: missing hydrogen atoms were added, the multiplicity of bonds, side chains of amino acid residues, networks of hydrogen bonds, desolvation energies were corrected. The method of flexible molecular docking implemented in the Glide program was used for Ligand-protein docking. The size of the docking area was consistent with the size of the ligand, the upper size limit was 15 Å. For each structure, up to 20 binding poses were generated. The optimality of the binding pose was determined based on the indicators GlideScore and Emodel, as well as on the clustering ability of docking solutions.

In the course of the study, it was confirmed that all compounds tend to bind at the same site located inside the transmembrane region. In addition, the features of the structure of the studied compounds allow each of them to be located in this site in an optimal manner with the formation of a number of hydrophobic interactions. The combination of the obtained data with the information on the ability of the compounds to inhibit MDM2 allowed us to identify five candidates that are promising multitarget agents capable of acting on two complementary cellular targets, MDM2 and P-glycoprotein. The use of such agents will significantly expand the possibilities of drug therapy for chemoresistant tumors.

The work was supported by the Russian Science Foundation (project No. 19-73-10150).

PHYTO4HEALTH – SECONDARY METABOLITES DATABASE OF RUSSIAN OFFICINAL MEDICINAL PLANTS

N. Ionov, D. Druzhilovskiy, D. Filimonov, A. Lagunin, V. Poroikov

Institute of Biomedical Chemistry, Moscow, Russia

Natural products (NP) are characterized by high structural diversity and a large set of biological activities. Thus, NP represent a promising source for the search for new pharmacological substances. The traditional use of Russian medicinal plants in medicine opens up prospects for discovering and developing new pharmaceuticals [1]. Evaluating natural products' phytochemical composition, extraction, purification of individual phytocomponents, and experimental testing of biological activity are time-consuming and expensive processes [2]. *In silico* methods based on analysis of «structure-activity» relationships are widely used to determine the most promising directions of natural products research. However, currently computational estimation of pharmacological potential of Russian medicinal plants gets complicated because there is no relevant database on phytocomponents in a suitable form for *in silico* analysis.

Our study aims to collect and aggregate data of chemical structures, properties, and experimental data on the interaction with molecular targets for phytocomponents of Russian officinal medicinal plants. Based on the data presented in GBIF database [3], we identified 233 species of Russian medicinal plants. Then, we extracted data on the phytochemical composition of those plants from 50 freely accessible databases and scientific publications. Data on molecular targets of the phytocomponents were compiled from scientific publications, PubChem, and ChEMBL databases. Data processing was performed using Python scripts.

As a result, we collected information on the phytochemical composition of 233 species of Russian officinal medicinal plants belonging to 71 families. We identified 3,128 unique phytocomponents whose structures are presented in MOL, InChI, InChi Key, and canonical SMILES formats, and supplemented with identifiers in PubChem and ChEMBL. We found data on phytocomponents presence in a separate part of the plant for 321 compounds. Thus, the total number of "Plant – Plant part - Phytocomponent" associations is 9,489. Interactions with 802 human molecular targets were found, and the total number of quantitative parameters characterized the interactions (IC50, Ki, Kd, EC50, AC50, ED50, GI50) is 13,688. The information in the database enriched by the results of the PASS Refined 2022 prediction at Pa>Pi threshold [4]. Number of predicted unique activities is 1954 and total number of records is 1332884.

The database with information on the phytocomponent composition of 233 Russian officinal medicinal plants Phyto4Health is freely available via Internet [5]. The collected data can be used to identify the previously unknown pharmacological potential of the 3,128 phytocomponents using *in silico* analysis.

The study was supported in the framework of the Program for Basic Research in the Russian Federation for a long-term period (2021-2030) (No. 122030100170-5)

1. Shikov, A.N., Narkevich, I.A., Akamova, A.V. et al. (2021) *Frontiers in Pharmacology*, 12, 1735.
2. Lagunin A. el al. (2014) *Natural Product Reports*, 31(11), 1585-1611.
3. URL: https://www.gbif.org/country/RU/summary, accessed on April 20th, 2022
4. Poroikov, V.V., Filimonov, D.A., Gloriozova, T.A. et al. (2019). *Russ Chem Bull*, 68(12), 2143–2154.
5. URL: http://way2drug.com/p4h, accessed on April 20th, 2022

PHENOLIC COMPOUNDS OF *VITEX NEGUNDO* AS PROMISING ANTIFUNGAL AND ANTIBIOFILM BIOTHERAPEUTIC CANDIDATE AGAINST *CANDIDA ALBICANS*

N. Jaiswal*, *A. Kumar

Department of Biotechnology, National Institute of Technology, Raipur, India

Natural products have been acting as an effective source of antimicrobial therapeutic agents for many years. However, drug discovery using natural products is a challenging task but if a drug is successfully developed to a plant product, it will be a boon towards better management of the fungal disease. because these products are safe and show low or tolerable. This research work was aimed to screen some phenolic compounds of a medicinal plant *Vitex negundo* to study the pharmacokinetics and ADMET for deciphering their antifungal potential which is unexplored yet. The pharmacokinetics has been determined by the PASS prediction tool, used to examine many parameters like drug-likeness, bioavailability, antifungal, and antibiofilm potency. The ADMET parameters were also determined. In our study 18 phenolic compounds from *V. negundo* has been detected with potent antifungal potential against *Candida* species (human fungal pathogen). Many compounds show a significant high efficiency against the biofilm of Candida which was quite interesting. Therefore, these compounds could be explored for drug development against Candidiasis in the future because natural products have shown beneficial uses. Results further indicate to process *V. negundo* phenolic compound in the development of effective biotherapeutics that could be used as new and promising antifungal (anti-planktonic and anti-biofilm) arsenals in the future.

TARGETING VIRAL PATHOGENS ASSOCIATED WITH MALARIA USING ANTIMALARIAL COMPOUNDS

A.K. Singh, A.N. Jha

Department of Molecular Biology and Biotechnology, Tezpur University, Tezpur, India

In malaria endemic countries, coinfections and cotransmissions of different viral pathogens are widely reported. Prior studies have shown that malaria can trigger the Epstein-Barr virus (EBV) reactivation in the body. Besides, the altered immunity due to malaria could increase susceptibility to acquire co-circulating viruses like SARS-CoV-2 or vice versa during pandemic times. The dual burden of pathogens can deteriorate health by inducing disease severity. There are no or limited antiviral therapies available against EBV and SARS-CoV-2. Exploring the novel antimalarials for checking antiviral efficacy and using them in such cases could be the efficient approach of 'hitting two birds with one stone'. We investigated the antiviral potency of medicine for a malaria venture's malaria box containing 400 drug-like or probe-like compounds with experimentally proven antimalarial activity. We utilized a molecular docking approach to screen these compounds against crucial proteins- EBNA1 of EBV and RdRp of SARS-CoV-2 respectively. Based on binding affinity we have shortlisted the top three compounds for each protein. Further, for validation of complex stability and binding, the protein-ligand complex is subjected to 100ns molecular dynamic simulation. All the compounds showed stable binding with respective proteins. Based on binding free energies, involvement of important residues from target sites, and ADMET properties of compounds, the top ligand for each protein is selected. Ligand B (MMV665879) for EBNA1 (ΔG_{bind}= -183.54 kJ/mol) and Ligand E (MMV665918) for RdRp (ΔG_{bind} = -172.23 kJ/mol) could act as potent inhibitors. These antimalarial compounds can hence be utilized for further experimental investigation as antivirals against EBV and SARS-CoV-2.

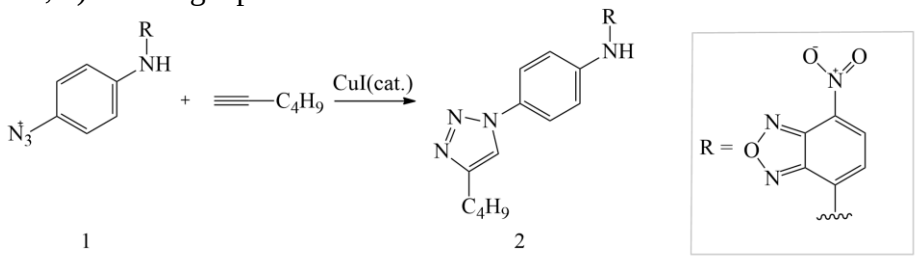
SYNTHESIS AND SOME PROPERTIES OF NBD-AZIDOANILINE

***V. Karpushenkova*[2], *Y. Faletrov*[1,2], *V. Shkumatov*[1,2]**

[1]*Research Institute for Physical Chemical Problems, Belarusian State University, Minsk, Belarus*

[2]*Faculty of Chemistry, Belarusian State University, Minsk, Belarus*

A conjugate of p-azidoaniline with 7-nitrobenzoxadiazol-4-yl chloride (NBD chloride) has been synthesized. The compound, NBD-azidoaniline (NBD-AzAn), seems to be new due to the absence of the structure in Pubchem database. Synthesis was performed at room temperature by mixing 4-azidoaniline with NBD-chloride in the presence of $NaHCO_3$ in the mixture of acetonitrile-methanol (2:1) followed by SiO_2 column chromatography similar to it was described [1]. Purity and molecular weight of NBD-AzAn was analyzed using TLC, electrospray-mass spectrometry, spectrofluorimetry and spectrophotometry. The presence of azide group in NBD-AzAn product were additionally confirmed using its known ability to undergo well-known "click" cycloaddition with alkynes. Synthesis of triazole (compound 2) was carried out by mixing NBD-AzAn (Scheme 1, **1**) with an excess of hexyne-1 using CuI as a catalyst followed by TLC analysis confirming formation of the triazole (Scheme 1, **2**) as a single product.



Scheme 1. Click modification of NBD-AzAn proving azide group in its structure and possibility for further functionalization using CuAAC

Additionally, changes in absorbance and fluorescence spectra in various solvents were estimated (absorbance maximum in methanol is at 478 nm). To evaluate biological properties of the compound we used an inverse high-throughput virtual screening using Autodock Vina [2] and a helper tool FYTdock [3]. 450 randomly chosen PDB structures of cytochromes P450 were used because the enzymes are known to reduce organic azides in hypoxic conditions [4] and our experience to use dockings to evaluate protein ligand interactions [5].

Docking results revealed 55 hits with binding energy values from -11.3 to -9.9 kcal/mol and complexes with structures PDB 6DWN, 6UDM (CYP1A1), 6CIZ, 6WW (CYP17), 3TDA (CYP2D6).

This study was supported in the framework of GPSR (Belarus) No. 20210560 and a personal grant of Wargaming.net to V.K.

1. Lomize, A.L., Pogozheva, I.D., Mosberg H.I. (2011) *J. Chem. Inf. Model.*, 51, 930.
2.Trott, O., Olson A.J. (2010) *J. Comp. Chem.*, 31, 455.
3.Faletrov, Y.V., Staravoitava, V.A., Dudko A.R. et al. (2022) PREPRINT (Version 1) available at Research Square [https://doi.org/10.21203/rs.3.rs-1456627/v1].
4. O'Connor L.J., Mistry I.N., Collins S.L. et al. (2017) *ACS Cent. Sci*., 3, 20.
5. Faletrov, Y.V., Karpushenkova, V.S., Zavalinich, V.A., Yakovets, P.S. (2021) *J. Belarusian State Univ. Chemistry*, 2, 25.

3D-STRUCTURE OF THE HUMAN REACTOME

A. Kozlova, T. Shkrigunov, A. Zgoda, A. Valishina, E. Sarygina, A. Lisitsa

Center for Scientific and Practical Education, Institute of Biomedical Chemistry, Moscow, Russia

The integration of genome, transcriptome, proteome, metabolome and metagenome data into representation of molecular interactions including small molecules and macromolecules with dynamic and kinetic information is the one of the central problems of systems biology. Protein-protein interactions (PPI) and metabolic pathways were revisioned to combine the metabolome-centered view of human reactome.

To create the first version of the 3D vision of the human reactome, we took raw stool metabolomic data received from 604 patients and deposited into The Inflammatory Bowel Disease Multi'omics Database (ibdmdb.org). Mass-spectra of samples have been obtained by HPLC-MS system (Nexera X2 U-HPLC systems (Shimadzu Scientific Instruments) coupled to Q Exactive/Exactive Plus orbitrap mass spectrometers (Thermo Fisher Scientific)) in .RAW format. The RAW-files were converted to mzML format using MSconvert and analyzed by the MetaboAnalystR - a R language package. For the found 542 metabolites, a search was made for proteins associated with them - enzymes and transporters - in the HMDB (hmdb.ca) and UniProt (www.uniprot.org). 3D-structures of the found proteins from the AlphaFold Database (alphafold.ebi.ac.uk) were also added. Information about post-translational modifications was taken from the iPTMnet database (research.bioinformatics.udel.edu). For each protein, splice variants were identified using the Ensembl (www.ensembl.org). The collected data are integrated with metabolic pathways by KeGG (www.genome.jp) and the relationship of target enzymes for pharmacologicaly active compounds with proteins of the signaling pathway cascades is discussed under this issue.

This work was financed by the Ministry of Science and Higher Education of the Russian Federation within the framework of state support for the creation and development of World-Class Research Centers "Digital biodesign and personalized healthcare" No. 75-15-2020-913.

BLOOD METABOLOME PROFILE OF OBESITY: PEERING TO THE LIPOPHILIC FRACTION

A. Kozlova, S. Tarbeeva, A. Lisitsa, E. Ilgisonis

Center for Scientific and Practical Education, Institute of Biomedical Chemistry, Moscow, Russia

Obesity is a multifactorial socially significant disease that can affect everyone regardless of gender, age and nationality. However, hereditary predisposition determines the probability of developing this disease. Despite the fact lifestyle affects obesity, few people are ready to admit having an improper eating behavior, that in the fast rhythm of life in megapolises is becoming widespread. Most people are not able to estimate the long-term consequences of unconscious actions imperceptibly turning into a habit. It is not a problem to diagnose obesity, but there are severe difficulties in establishing recommendations without significant changes in lifestyle and behavior patterns.

Metabolomics is facilitated to uncover mentioned problem using blood plasma metabolome like a mirror of eating behavior, lifestyle, as well as hereditary predisposition and concomitant latent metabolic disorders. Most of the metabolome can be detected by high-throughput technology - direct injection mass spectrometry (DIMS) with electrospray ionization. However, under DIMS method some lipophilic metabolites remain invisible. Gas chromatography coupled with mass spectrometry (GC-MS) allows to determine the majority of lipophilic compounds due to the separation of compounds in the sample into fractions. Additional separation of fractions in the GCxGC-MS method increases the information content of the output data: thus, more compounds can be detected.

For an orthogonal analysis of the blood plasma samples, we compared the data obtained by DIMS and GCxGC-MS methods. Reusing bioinformatic pipelines we have expanded the understanding of metabolome of the obese patients.

The study was performed employing "Avogadro" large-scale research facilities, and was financially supported by the Ministry of Education and Science of the Russian Federation, Agreement No. 075-15-2021-933, unique project ID: RF00121X0004.

ACYLATION MECHANISM OF PENICILLIN-BINDING PROTEINS 2 FROM STRAINS FA19, 35/02 AND H041 BY CEFTRIAXONE

***A. Krivitskaya*[1], *M. Khrenova*[1,2]**

[1]*The Federal Research Centre "Fundamentals of Biotechnology" of the Russian Academy of Sciences, Moscow, Russia*

[2]*Lomonosov Moscow State University, Moscow, Russia*

Resistance to antibacterial drugs increases rapidly. Resistant strains of the causative agent of gonorrhea, the bacterium *Neisseria gonorrhoeae*, to extended-spectrum cephalosporins have been repeatedly identified in clinical practice [1]. Ceftriaxone antibiotic, which is frequently used in the treatment of the disease, is among them. It is known that resistance arises due to mutations in the allele *penA* gene encoding *N. gonorrhoeae* penicillin-binding protein 2 (PBP2). PBP2 has transpeptidase activity and catalyzes the last stages of peptidoglycan formation in the bacterial cell wall, carrying out the transpeptidation reaction [2]. The natural substrate of PBP2 is structurally similar to β-lactam antibiotics. Therefore, PBP2 is able to bind ceftriaxone, which blocks its transpeptidase activity, which is vital for the bacterium. The antibacterial effect of ceftriaxone against *N. gonorrhoeae* is based on this.

The PBP2s from the wild strain FA19 and mutated PBP2 from strains 35/02 and H041 are known. The difference in k_2/K_s for the reaction between ceftriaxone and PBP2 from 35/02 and H041 in relation to the FA19 is 151 and 2307 times, respectively. Previous studies determined the mutation influence on rotation and mobility to β_3-β_4 loop [3]. It is supposed that the mobility of this loop can play a key role in affinity, acylation or deacylation or in the several of these processes. Our study aims at establishing the mechanisms of acylation of PBP2s from FA19, 35/02 and H041 strains to identify the mechanistic reasons of appearance of resistance by methods of computational chemistry. In the course of our work, we have analyzed the enzyme-substrate complexes of PBP2 with ceftriaxone and identified reactive and non-reactive complexes. We determined the acylation mechanisms of PBP2s by ceftriaxone. We also considered the substrate binding to the PBP2s active site and analyzed the conformation changes in the β_3-β_4 loop.

We demonstrate that the β_3-β_4 loop is more labile in FA19, more stable in 35/02 and practically doesn't move in H041. This difference affects substrate binding and active site formation for these strains. The enzyme-substrate complexes are different for FA19 and mutated strains. This defines different acylation mechanisms for wild type and mutated strains. Acylation reaction occurs in three elementary steps in FA19 with formation of the first stable intermediate. While in mutant strains the reaction occurs in two steps. Thus, we determined the acylation mechanism of the PBP2 from the wild type FA19 and mutant strains, 35/02 and H041, and also evaluated the effect of mutations on the mobility of the β_3-β_4 loop. These theoretical results can serve as a reliable basis for the development of inhibitors of this process.

This study was supported by Russian Science Foundation (grant No. 18-74-10056).

1. Sauvage, E., Kerff, F., Terrak, M. et al. (2008) *FEMS microbiology reviews*, 32, 234.
2. Tomberg, J., Fedarovich, A., Vincent, L. R. et al. (2017) *Biochemistry*, 56, 1140.
3. Tomberg, J, Unemo, M., Ohnishi, M. et al. (2013) *Antimicrobial agents and chemotherapy*, 57, 3029.

COMPUTER ANALYSIS OF THE BIOLOGICAL ACTIVITY OF HYDROXY DERIVATIVES OF HYDROQUINOLINES AND ASSESSMENT OF THE PROTECTIVE POTENTIAL OF 6-HYDROXY-2,2,4-TRIMETHYL-1,2-DIHYDROQUINOLINE IN A RAT MODEL OF DRUG-INDUCED LIVER INJURY

***E.D. Kryl'skii*[1], *S.E. Popova*[1], *T.N. Popova*[1], *Kh.S. Shikhaliev*[1], *L.V. Matasova*[1], *A.I. Parkhomenko*[1], *V.V. Poroikov*[2]**

[1]*Voronezh State University, Voronezh, Russia;*
[2]*Institute of Biomedical Chemistry, Moscow, Russia*

Acetaminophen is currently the most widely used analgesic and antipyretic. At high doses, this drug can cause severe liver damage. Oxidation of acetaminophen by cytochrome P450 enzymes produces N-acetyl-p-benzoquinonimine, a toxic metabolite that binds reduced glutathione. Depletion of the glutathione pool leads to the accumulation of reactive oxygen species (ROS), which cause oxidative stress leading to hepatocellular necrosis and apoptosis [1].

Despite the large number of hepatoprotective agents on the pharmaceutical market, they all have a number of disadvantages, therefore, the search for new potential drug precursors is an urgent task. Hydroxy derivatives of hydroquinolines are of interest in this respect. One of the representatives of this group of compounds is the synthetic antioxidant ethoxyquin (6-ethoxy-2,2,4-trimethyl-1,2-dihydroquinoline) used for the prevention of liver diseases in pigs [2]. Nevertheless, due to the presence of several negative effects of this compound [3], its use as a hepatoprotective agent in humans was inexpedient. Thus, we had the task to screen the biological activity of ethoxyquin derivatives in order to find a more suitable drug candidate.

Given the large variety of chemical structures available for synthesis, computer prediction of their biological activity seems to be the preferred method for selecting the most promising molecules. Using PASS Online software, we revealed that 6-hydroxy-2,2,4-trimethyl-1,2-dihydroquinoline (DHQ) has the highest probability of showing protective properties: antioxidant activity, inhibition of lipid peroxide oxidation, membrane integrity agonist, anti-inflammatory activity, etc. To study the protective activity of DHQ *in vivo*, we used a rat model of acetaminophen-induced liver damage. The results showed that DHQ administration to rats with pathology reduced the activity of marker enzymes of hepatocyte cytolysis, in particular alanine aminotransferase, aspartate aminotransferase, and gamma-glutamyltranspeptidase. Apparently, the main mechanism of DHQ's protective activity was its ability to inhibit ROS-induced oxidation. Thus, we showed that the tested compound normalized the parameters of biochemiluminescence reflecting the intensity of oxidative processes and total antioxidant activity, reduced the level of primary products of lipid peroxidation, and increased the activity of aconitate hydratase, which is a sensitive target of ROS. In addition, DHQ promoted a change in the activities of superoxide dismutase, catalase, glutathione peroxidase, glutathione reductase, and glutathione transferase toward control values in animals with pathology.

The data obtained testify to the relevance of the development of new hepatoprotective agents with antioxidant activity on the basis of DHQ. In addition, computer screening of the biological activity of hydroxy derivatives of hydroquinolines allowed us to identify a new group of substances of interest as potential drug precursors.

This study was supported by RFBR grant, project No. 20-04-00526.

1. Ulger, O., Kubat, G.B., Cicek, Z. et al. (2021) *Life Sciences*, 279, 119669.
2. Bagautdinov, A.M., Baymatov, V., Farkhutdinov, R.R. et al. (2013) Patent RU2483695 C1, Bulletin № 16.
3. Blaszczyk, A. (2006) *Toxicology Letters*, 163, 77-83.

COMPUTATIONAL SCREENING OF NATURAL INHIBITORS FOR POLYCYSTIC OVARY SYNDROME THROUGH POLYPHARMACOLOGY APPROACH

H.F. Escobar-Morreale

Department of Endocrinology and Nutrition, Hospital Universitario Ramón y Cajal, Madrid, Spain

One of the most common metabolic and endocrine hormonal disorder in young women is Polycystic ovary syndrome. Around the world, 4%-20% of women of reproductive age are affected by this disease. In the absence of a specific diagnosis, PCOS exhibits a wide range of symptoms of androgen overload and ovarian dysfunction. PCOS is a complex multiple genetic condition with strong epigenetic and environmental consequences, as well as nutritional and lifestyle limitations. Diet is vital in PCOS; a good diet should be incorporated. Various sorts of dietary sources should be consumed during PCOS. These dietary sources include a variety of metabolites, some of which are active metabolites and others which are inactive metabolites. Molecular descriptors provide information about the activity of metabolites. They help reduce the dimensionality produced by a virtual screening approach. Furthermore, they aim to sample huge datasets, analyze the process statistically, and build a screening filter to discover compounds with polypharmacological or promiscuous activity. PaDel Descriptors, a molecular descriptor generating software, is used to construct descriptors in this study.

Later, the principal component analysis (PCA) method is used to select the most significant chemical and biological molecular descriptors. For lead identification, the substances found in dietary sources are summarized from the literature in order to identify active molecules and possible targets. Genes are established as the intersection of PCOS-associated genes and the expected targets of active compounds, which aided in the formation of a pharmacological network. For target identification, comparing patients with PCOS to non-PCOS controls, differential expression analysis was used to determine which genes were differentially expressed. Finally, important genes are identified at the junction of DEGs and target genes. Ultimately, docking studies for both lead and target helps in determine the molecular interactions and binding affinities of the inhibitors with the associated targets, deciphering the regulatory mechanisms and performing molecular assays for future studies.

1. Escobar-Morreale., H.F. (2018). *Nature Reviews Endocrinology*, 14(5), 270-284.

PERSPECTIVES OF 15-LIPOXYGENASE INHIBITOR ANALOGS AS POTENTIAL ANTI-INFLAMMATORY AGENTS

V.R. Khairullina[2], Y.Z. Martynova[1], A.R. Gimadieva[2], A.G. Mustafin[1,2]

[1]Department of Chemistry, Bashkir State University, Ufa, Russia;
[2]Ufa Institute of Chemistry of the Russian Academy of Sciences, Ufa, Russia

Lipoxygenases are a family of iron-containing waste enzymes that catalyze the formation of leukotrienes from arachidonic acid. Lipoxygenase (LOX) products are important mediators of the pathophysiology of asthma, heart disease, and cancer. The aim of this work was to model some nitrogen-, oxygen-, and sulfur-containing heterocyclic compounds with pronounced inhibitory activity against 15-lipoxyoxygenase [1], and to build QSAR (Quantitative Structure–Property Relationships) models based on them. The studies were carried out using the GUSAR 2019 program (General Unrestricted Structure Activity Relationships) based on two training and two test sets according to the method described in [2]. As a result, six statistically significant stable consensus models of QSAR predict the numeral values of IC_{50} for LOX inhibitors (Table 1). They are applicable for virtual screening and search for new compounds. In addition, structural descriptors have identified that regulation of the activity of LOX inhibitors is allowed.

Table 1. Characteristics and prediction accuracy of IC_{50} values for consensus-models M1-M6

Training set	Models	N	R^2_{TrS}	R^2_{TS}	F	S.D.	Q^2	V
QSAR model based on QNA-descriptors								
TrS1	M1	84	0.962	0.800	13.026	0.443	0.800	17
TrS2	M4	86	0.954	0.875	9.851	0.486	0.758	18
QSAR model based on MNA-descriptors								
TrS1	M2	84	0.955	0.794	10.365	0.486	0.760	16
TrS2	M5	86	0.954	0.861	8.242	0.511	0.734	18
QSAR model based on QNA- and MNA-descriptors								
TrS1	M3	84	0.969	0.755	13.180	0.420	0.822	18
TrS2	M6	86	0.965	0.836	11.486	0.440	0.807	19

N – number of structuresinthe training set; R^2_{TrS} - a multiple coefficient of determination calculated for compounds from the training set; R^2_{TS} - a multiple coefficient of determination calculated for compounds from the test set; Q^2 – a cross-validated R^2 calculated during leave-one-out cross-validation procedure on data of the training set; F – Fisher's coefficient; SD –standard deviation; V- the number of variablesin the finalregression equation

The reported study was funded by the Russian Science Foundation for Basic Research according to the research project No. 19-73-20073.

1. Pontiki, E., Hadjipavlou-Litina, D. (2008) *Med. Res. Rev.*, 28, 39-117.
2. Khairullina, V., Safarova, I., Sharipova, G., Martynova, Y., Gerchikov, A. (2021) *Molecules*, 26, 421.

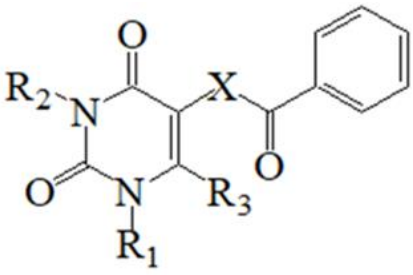
MOLECULAR DOCKING OF URACIL DERIVATIVES INTO THE ACTIVE CENTER OF DIHYDROPYRIMIDINE DEHYDROGENASE

V.R. Khairullina[2], Y.Z. Martynova[1], A.G. Mustafin[1,2]

[1]Department of Chemistry, Bashkir State University, Ufa, Russia;
[2]Ufa Institute of Chemistry of the Russian Academy of Sciences, Ufa, Russia

The steric complementarity of 20 uracil A1-A20 derivatives synthesized in 2020 with the active center of pig liver dihydropyrimidine dehydrogenase (DPYD) was studied. A macromolecule with the code 1gth (chains A, B) was chosen from the PDB as a DPYD model [1]. The positioning of ligands in the DPYD active center, as well as the E_{bind} energy, was determined by molecular docking using the AutoDock Vina program [1-2]. The preparation of ligand and protein structures was carried out in the AutoDockTools program. The active center of the macromolecule was placed in a three-dimensional box with a size of 22 Å. The position of the native DPYD ―5-ioduracil ligand was taken as the center of the box [1].

It has been established that almost all ligands are not sterically complementary with the active center of DPYD, in which thymine metabolism occurs. However, compounds A1-A4, A6, A7, A11-A18 are located in the pool at the entrance to the active center of this protein (Fig. 1), while their binding energy to E_{bind} is comparable to the similar characteristic for uracil, thymine and 5-fluorouracil (the last three ligands are located in the active center of DPYD). This fact indicates that it is impossible to exclude the possibility of inhibition of DPYD activity in vivo by compounds A1-A4, A6, A7, A11-A18, and also suggests that the co-administration of these compounds with 5-fluorouracil as part of antitumor drugs does not exclude the likelihood of a decrease in the rate of biodegradation of 5-fluorouracil, and consequently, the probability of reducing its therapeutic dose. Compounds A5, A8-A10, A19, and A20 are not sterically complementary to either the active center of DPYD or the pool at its entrance and, therefore, with a high probability will not inhibit the activity of the DPYD enzyme in vivo. The potentially bioactive conformations of compounds A1-A4, A6, A7, A11-A18 in the DPYD pool and the factors of stabilization of their position in it were determined. In particular, mainly hydrogen bonds with amino acid residues Thr 761 (chain A), Arg776 (chain B), Gly933 (chain B), Thr934 (chain B), Gly674 (chain A) stabilize the position of uracil molecules in the pool at the entrance to the active center of pig liver DPYD.



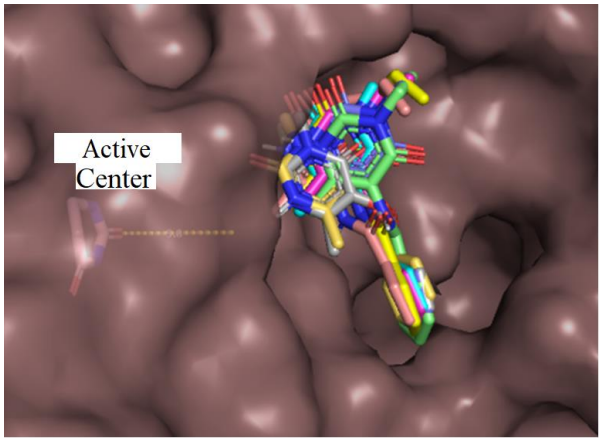


Figure 1. Results of positioning of uracil A1-A4, A6, A8, A11, A12 derivatives synthesized in 2020 in the active center of DPYD from pig liver

The reported study was funded by the Russian Science Foundation for Basic Research according to the research project No. 19-73-20073.

1. URL: https://www.rcsb.org/
2. URL: http://vina.scripps.edu/

COMPUTATIONAL PREDICTION OF SUSCEPTIBILITY TO LIMITED PROTEOLYSIS FOR PROTEINS WITH KNOWN 3D STRUCTURE

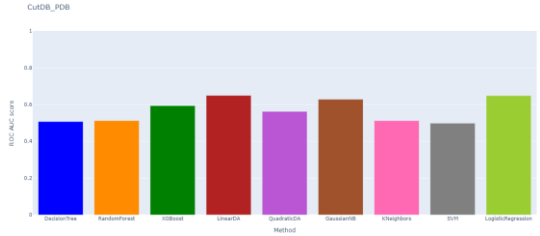
***E.V. Matveev*[1], *M.D. Kazanov*[1,2]**

[1]*Institute for Information Transmission Problems (Kharkevich Institute, Russian Academy of Science)*
[2]*Skolkovo Institute of Science and Technology*

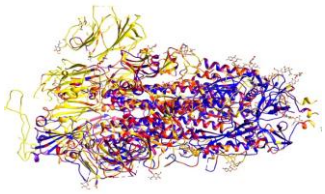
Identification of the protease substrates is important for elucidating mechanisms of many molecular processes in the living cell, including apoptosis, cell proliferation, protein activation, or degradation. Computational predictions of proteolytic events can substantially reduce the amount of experimental work required to identify protease substrates. The majority of existing computational methods for the protease cleavage site prediction broadly uses known primary protease specificity but is limited in using the structural data of potential substrates. However, the susceptibility of protein substrates to limited proteolysis is known to depend strongly on the structural context of the cleaved peptide bond. Moreover, the databases containing structural information are abundant especially given the recent breakthrough in predicting protein's 3D structure, which opens up new possibilities for biological prediction models. To our knowledge, there is only one method, that uses 3D structures of potential protease substrates, and there are no methods for characterizing the susceptibility of protein regions to proteolytic processing.

Earlier, we comprised curated training datasets of proteolytic events obtained from CutDB (~600 proteolytic events) and MEROPS databases (~7000 proteolytic events) and constructed a structural method for predicting the susceptibility of protein's peptide bonds to proteolytic processing based on machine learning approach. Here, we used data from the experiment on proteolytic processing of E. coli proteins, whose structures are mostly known, by MMP9 and MMP25 proteases to obtain data both on the sequence and structural preferences of particular proteases. We modeled sequence preference of proteases using Positional-Weight Matrices and obtained sequence and structural scores for each peptide bond of considered substrates. Analysis of the sequence-structural score plots shown that cleaved sites with high structural scores can possess both moderate and high sequence scores, while cleaved sites with low structural scores usually have high sequence scores. This analysis will allow us to build a combined structural-sequence model with predicting the susceptibility of substrate's peptide bonds to limited proteolysis.

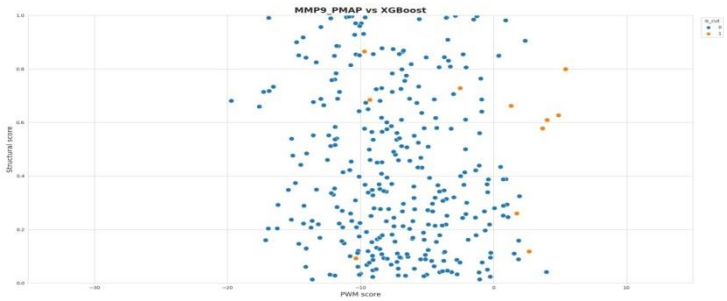
The reported study was funded by RFBR according to the research project No. 20-04-60066.



I. ROC AUC scores of structural models trained on CutDB dataset.



II. Visualisation of the structural susceptibility to proteolysis.



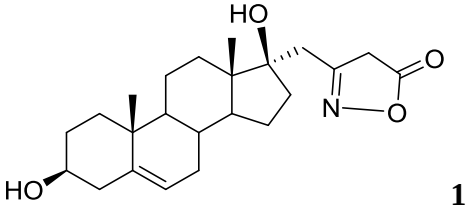
III. The correlation between structural and sequence (PWM) scores for 1VLY_A structure. PWM model was constructed for MMP9 protease. The structural model was built using XGBoost ML algorithm. Known cleaved peptide bonds are presented as orange points.

STUDY OF THE BIOLOGICAL ACTIVITY OF 21-NORSTEROID 20-AZOLES - POTENTIAL LIGANDS OF RECEPTORS CONTAINING A STEROL-SENSITIVE DOMAIN (SSD)

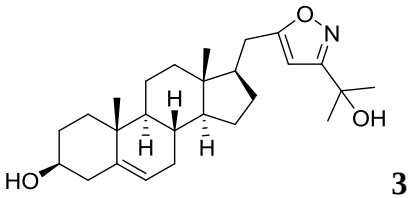
A. Aleksandrova[2], A. Mekhtiev[1], K. Shcherbakov[1], O. Timoshenko[1], A. Gilep[3], A. Rudovich[3], V. Zhabinskii[3], V. Khripach[3]

[1]Institute of Biomedical Chemistry, Moscow, Russia
[2]Pirogov Russian National Research Medical University, Moscow, Russia
[3]Institute of Bioorganic Chemistry NASB, Minsk, Belarus

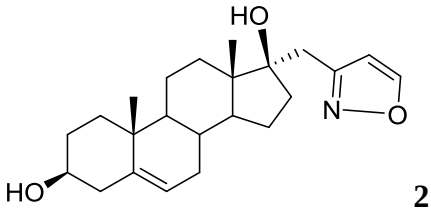
Hedgehog (Hh) signaling is activated in embryonic cells, but the proliferation of many tumor lines is regulated by Hh proteins, as well as its activation is associated with liver fibrosis and degeneration. A personalized approach to solving this socially significant problem allows not only to develop new ways of diagnosing tumor diseases, but also to predict and create new potential pharmaceutical agents for targeted therapy of tumors and liver fibrosis. Personalized medicine and targeted therapy play a special role in the diagnosis and treatment of oncological diseases, because the same types of tumors in different patients may have different genesis, and therefore require an individual approach to their diagnosis and treatment. The development of a personalized algorithm for identifying ligands of target proteins for new pharmaceutical agents based on the study of their biological activity would solve the problem of choosing the optimal therapy and bring the strategy of cancer treatment to a qualitatively new level. Our work is aimed at studying the biological activity of new steroid compounds interacting with the sterol-sensitive domain of various proteins involved in the regulation of tumor cell proliferation. Steroid molecules are not only antagonists of these proteins, but also participate, e.g., as cholesterol, in signal transmission in the molecular cascade. The study of ligand-protein interactions allows to select the right compounds for targeting.



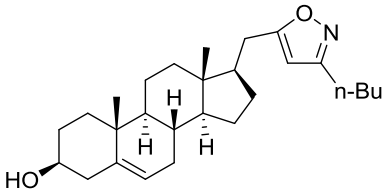
1



3



2



4

Two series of 21-norsteroid 20-azoles were tested on various models of cell lines and reconstructed systems *in vitro* with the leading compounds **1-4**. Docking and computer modeling of azoles **1** and **2** in the steroid binding site of the ligand-binding domain of the androgen receptor (AR) have shown that both compounds can be AR antagonists. However, the results of spectral titration of human CYP17A1 with synthesized compounds and positive control, abiraterone, showed that the studied compounds do not inhibit CYP17A1. Studies on cell lines have shown that LNCaP prostate carcinoma cells are characterized by growth suppression of IC_{50}=40 μm and 20 μM, respectively, with an induction of apoptosis of about 30%. Compounds **3** and **4** were studied on Hela, A549, HCT116, PC3, and HepG2 cell lines. In all tumor cell lines, the studied substances showed high cytotoxicity and blocked the growth of tumor cells for long periods of incubation. We assume that the action of the compounds is associated with the suppression of Hh signaling activated in these cells.

This work was supported by the Ministry of Science and High Education of the Russian Federation within the framework of state support for the creation and development of World-Class Research Centers "Digital Biodesign and Personalized Healthcare" (No 75-15-2022-305) and by the Belarusian Foundation for Fundamental Research (project X22Mldg-001).

2-SUBSTITUTED ALLYL BROMIDES AS ELECTROPHILIC AND NUCLEOPHILIC BUILDING BLOCKS. SYNTHESIS OF NEW HETEROCYCLIC COMPOUNDS AND *IN SILICO* ANALYSIS OF BIOACTIVITY OF OBTAINED COMPOUNDS

***I.V. Mineyeva*[1], *V.M. Shkumatov*[1,2]**

[1] *Faculty of Chemistry, Belarusian State University, Leningradskaya str. 14, Minsk, Belarus*
[2] *Research Institute for Physical Chemical Problems, Belarusian State University, Leningradskaya str. 14, Minsk, Belarus*

For the first time, based on 2-substituted functionalized allyl bromides, an effective method was developed for the preparation of various benzo[f]coumarins **8-15**, allylated 3,4-dihydropyrimidine-2(1H)-thiones **16-23**, isoxazolines **24-28**, oxazoles and pyrazoles **25- 28**.

The analysis of pharmacological activity was carried out using computer technologies in silico on the PassOnline platform. (http://way2drug.com/passonline/predict.php). Possible effective interactions with molecular targets are calculated and predictions of various types of activity are made. A prediction of interaction with tumor and non-tumor cell lines has been made. All studied substances exhibit antibacterial, antifungal, antiviral activity.

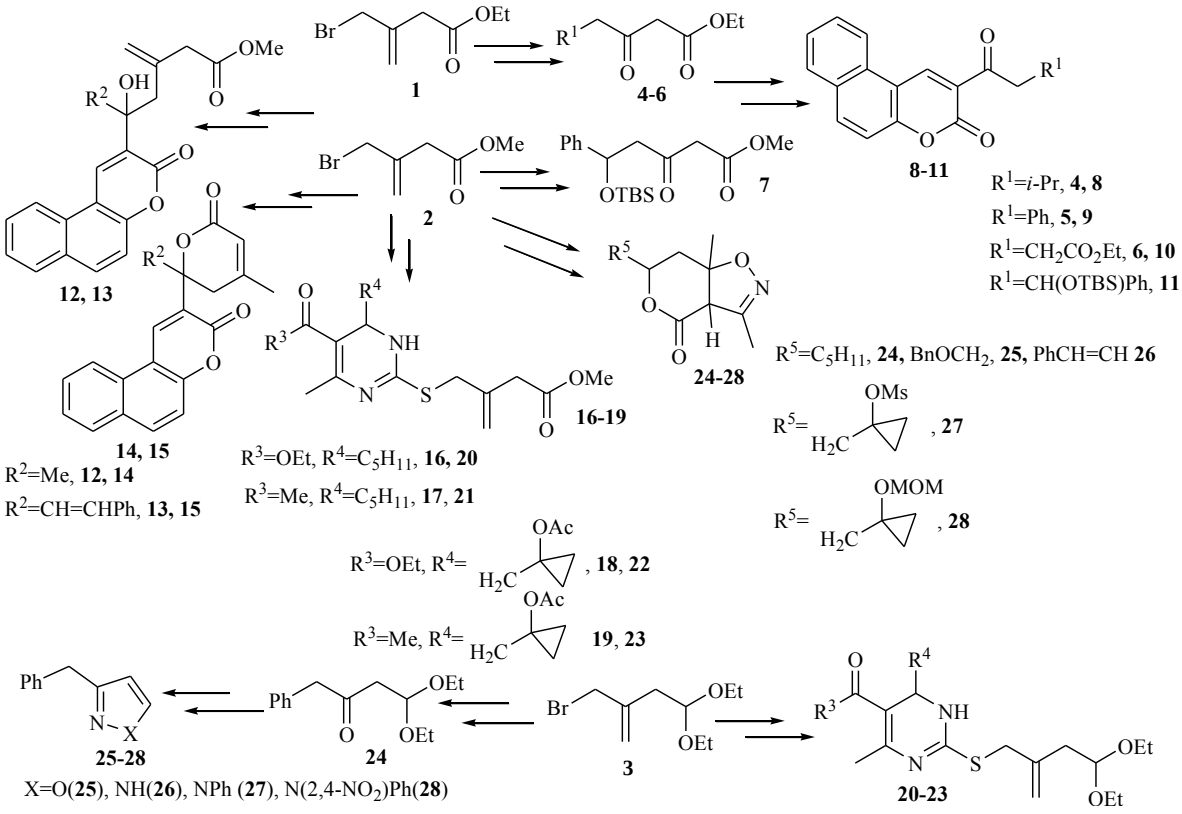


Figure 1. Structures of all new heterocycles under consideration

The results obtained indicate the possibility of obtaining new heterocyclic compounds based on 2-substituted allyl bromides and evaluating their biological properties by in silico methods for the purpose of obtaining new molecular tools and drug prototypes.

This work was supported by the SPSR (Belarus) No. 20211462.

DOCKING OF ORIGINAL THIAZOLO[3,2-A]PYRIMIDINES WITH SARS-COV-2 PROTEINS

***I.V. Mineyeva*[1], *Y.V. Faletrov*[1,2], *V.A. Staravoitava*[1,2], *V.M. Shkumatov*[1,2]**

[1] *Faculty of Chemistry, Belarusian State University, Leningradskaya str. 14, Minsk, Belarus*
[2] *Research Institute for Physical Chemical Problems, Belarusian State University, Leningradskaya str. 14, Minsk, Belarus*

The purpose of this work is to evaluate the previously synthesized compounds [1] for activity against SARS-CoV-2 using molecular docking. Molecular docking of 452 SARS-CoV-2 proteins and 21 synthesized compounds was performed using the FYTDOK tool, which employs the AutoDock Vina, Binana, oBabel executable files and python scripts [2]. Consequently, nearly 9500 protein-ligand complexes with binding energies ranging from -3.6 to -9.9 were obtained. The largest number of complexes with different SARS-CoV-2 proteins structures, possessing binding energies below -7.5, were obtained for compounds **1**, **2**, and **3**. Compound **1** showed 64 such complexes with different PDB structures of SARS-CoV-2 enzymes, among them 19 complexes with different structures of NSP13 protein (E_{bind} =-8.1 – -7.5), 17 with NSP16 protein (E_{bind} =-8.1 – -7.5) and 6 with NSP3 macrodomain (E_{bind} =-9 – -7.9). For compound **2** were found 145 complexes with different SARS-CoV-2 structures, among them 51 complexes with different structures of NSP13 protein (E_{bind} =-8.4 – -7.5), 6 with NSP3 macrodomain (E_{bind} =-9.7 – -7.7) and 36 with 3CL MPro (E_{bind} =-8.2 – -7.5). Compound **3** demonstrated 118 complexes with different SARS-CoV-2 structures, among them 52 complexes with different structures of NSP13 protein (E_{bind} =-8.3 – -7.6), 6 with NSP3 macrodomain (E_{bind} =-9.9 – -8.3) and 18 with NSP16 protein (E_{bind} =-8.5 – -7.7). Moreover, acute rat toxicity prediction was performed for these three compounds using GUSAR software [3]. The assay indicated that all three substances belong to the 4th class of toxicity (practically non-toxic and not an irritant), as well as remdesivir and atanazavir, known antivirals. Compound **1** has the best LD50 values among the three ligands: 56.570 mg/kg for the intravenous route and 1175 mg/kg for the oral route.

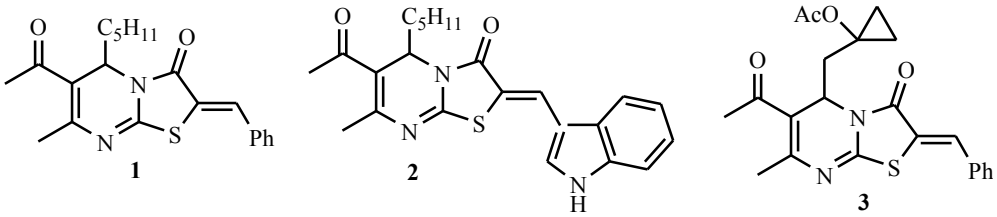


Figure 1. Structures of thiazolo[3,2-a]pyrimidines which demonstrates the best affinity to target structures of SARS-CoV-2 enzymes under consideration

This study was supported by Belarusian grants BRFFR X21COVID-030 and GPFR No. 20210560.

1. Mineeva, I.V. Faletrov, Y.V., Staravoitava, V.A., Shkumatov, V.M. MedChem-Russia 2021. DOI: 10.19163/MedChemRussia2021-2021-426
2. Faletrov, Y.V., Staravoitava, V.A., Dudko A.R. et al. (2022) PREPRINT (Version 1) DOI: 10.21203/rs.3.rs-1456627/v1
3. Lagunin A., Zakharov A., Filimonov D., Poroikov V. (2011). *Molecular Informatics*, 30(2-3), 241–250.

SYNTHESIS OF NEW 1,4-DIHYDROPYRIDINES AND *IN SILICO* ANALYSIS OF BIOACTIVITY OF COMPOUNDS BY THE METHOD OF MOLECULAR DOCING

I.V. Mineyeva [1] **, *Y.V. Faletrov*** [1,2] **, *V.A. Staravoitava*** [1,2] **, *V.M. Shkumatov*** [1,2]

[1] *Faculty of Chemistry, Belarusian State University, Minsk, Belarus*
[2] *Research Institute for Physical Chemical Problems, Belarusian State University, Minsk, Belarus*

For the first time, an efficient method was developed for the preparation of 1,4-dihydropyridines (**1**-**7**) containing a residue of aliphatic and β-hydroxycyclopropane aldehydes using europium(III) chloride hexahydrate as a catalyst for the multicomponent Hantzsch reaction. The biological properties of the obtained compounds were evaluated by modeling the permeability through the phospholipid bilayer and molecular docking in relation to protein kinases and human cytochromes P450.

A theoretical assessment was made of the penetration of the studied substances into the cell by the efficiency of their passive diffusion through the lipid bilayer. The evaluation was carried out using the PerMM service, which allows predicting the membrane permeability for passive diffusion of this molecule based on the 3D structure of the studied molecule. The logarithms of the permeability coefficients for models of three different membranes have a value greater than -4.35, therefore, all the studied derivatives are able to penetrate through the cell membrane and participate in intracellular regulation. The highest permeability was predicted for compound **2**.

For the first time, dihydropyridines were docked to a pool of human protein kinases and cytochromes P450 using the Autodock Vina program. Compounds **4** and **7** have the potential for experimental testing of kinases taken for calculation, since they showed *in silico* affinity for these target proteins comparable to or greater than leader compound **8**. Calculations for human P450 cytochromes showed that the range of Ebind values was from 10.1 to 6.2 kcal/mol, and dihydropyridine **4** showed better interaction parameters than the known dihydropyridine **8** with anticancer activity. Good affinity (Ebind values from 9.5 kcal/mol and below) has been shown for *in silico* interaction with important steroid-converting cytochromes P450 (CYP11A1, CYP46A1, CYP51, CYP19A1), as well as P450 involved in drug metabolism (CYP3A4).

The analysis of pharmacological activity was carried out *in silico* on the PassOnline platform and revealed about 30 most probable types of pharmacological activity (http://way2drug.com/passonline/predict.php).

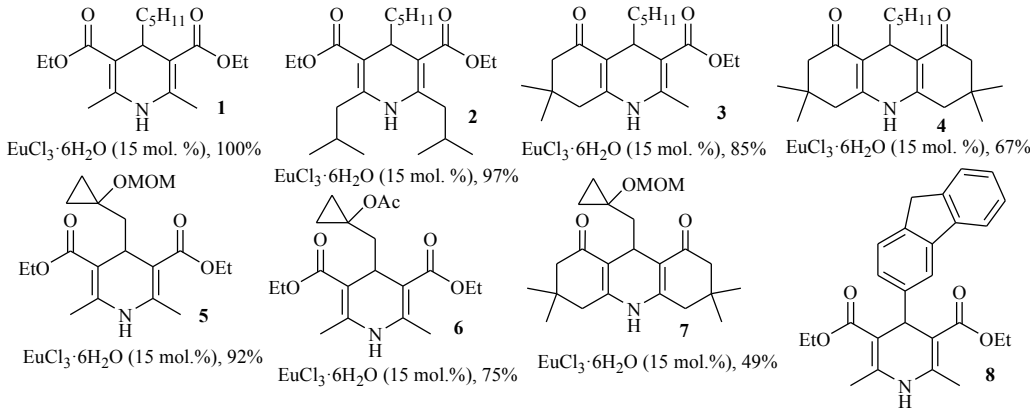


Figure 1. Structures and yield of all new 1,4-dihydropyridines

The results obtained indicate the possibility of obtaining new aliphatic derivatives of Hantzsch esters and evaluating their biological properties by *in silico* methods for the purpose of obtaining new molecular tools and drug prototypes and enables effective treatment selection for clinically complicated cases with no obvious treatment options

This work was supported by the SPSR (Belarus) No. 20210560.

A COMPREHENSIVE COMPUTATIONAL PHARMACOKINETICS IDENTIFICATION OF BIOTRANSFORMED LEADS FROM *CURCUMA CAESIA* ROXB

***K.S. Mukunthan*[1], *S. Balaji S*[1], *N.P. Trupti*[2]**

[1]*Department of Biotechnology, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, India*

[2]*Department of Integrative Biology, Vellore Institute of Technology, Vellore, India*

The biotransformation of drug happens with various metabolic stages and most of the investigations are focused on metabolizing enzymes such as CYP because it is the entry point of interaction for any drug candidate that enters the liver. Most predictions of ADMET, consider the inhibition of Cytochrome P450 with lead structures at the atomic level with the models of human Cytochrome P450. To date, predictions of monooxygenase based chemical transformation of the drug candidates with the metabolizing enzymes are still in primitive levels of research. In the current study, *in silico* screening [103 (native compounds) + 103 (Reduced form) + 103 (oxidized form)] of 309 exclusive compounds of *C.caesia* was carried to identify lead molecule. *In silico* absorption, distribution, novel metabolic prediction and toxicity resulted in the discovery of lead molecule i.e. Compound No 3. Further molecular docking assessment of the metabolically biotransformed leads identified using simplified pharmacophore analysis showed good inhibition potential against Peptidyl-prolyl cis-trans isomerase (PIN) and thus promises to be a good cancer inhibitor. The cost-effective approach presented in this paper could be used to filter toxic compounds from the drug discovery cycle.

1. Zou, Y., Li, H., Graham, E.T. et al. (2020) *Nat. Chem. Biol*., 16, 302–309.
2. Lee, S. K., Lee, I.H., Kim, H.J., Chang, G.S., Chang, J.E. (2003) The PreADME Approach:web-based program for rapid prediction of physic-chemical, drug absorption and drug-like properties. Blackwel Publishing, Boston, MA, 418-420. URL: http://preadmet.bmdrc.org

PAN-CANCER GENE EXPRESSION ANALYSIS OF CHOROMBOX GENE FAMILY REVEALED OVEREXPRESSION OF CBX2 GENE IN MULTIPLE CANCERS

***A.A.T. Naqvi*[1], *S.A.M. Rizvi*[1], *M.I. Hassan*[2]**

[1]*Department of Computer Science, Jamia Millia Islamia, New Delhi, India*
[2]*Centre for Interdisciplinary Research in Basic Sciences, Jamia Millia Islamia, New Delhi, India*

The chromobox (CBX) gene family is an essential part of the canonical polycomb repressive complex 1 (PRC1) involved in chromatin remodeling. Chromatin remodeling is a key event in the transcriptional regulation of various genes, making them significant for various developmental pathways. Disturbed chromatin structure is linked with the onset and progression of cancers. Altered expression of CBX genes is directly associated with abnormalities in chromatin modulation, hence playing a crucial role in carcinogenesis. CBX2 is a key member of the CBX gene family, which has a role in various regulatory pathways. Therefore, we performed a pan-cancer expression analysis of CBX genes in 18 cancer datasets obtained from The Cancer Genome Atlas (TCGA) repository. The normalized counts of mRNA expression levels were log2 transformed prior to expression analysis. The expression levels were measured in the tumor samples and compared with their normal adjacent samples to evaluate the differential expression of these genes. Receiver operating characteristic (ROC) analysis was also performed, and area under curve (AUC) values were calculated to assess the statistical significance of differential expression. Fold change (FC) values of more than 1.5 were considered significant for differential expression. The results of differential expression analysis and AUROC revealed elevated expression of CBX2 in multiple cancers. The highest level of CBX2 overexpression was observed in rectum adenocarcinoma with an FC value of 1.98 and an AUC of 0.99, followed by glioblastoma multiforme (FC=1.91 and AUC=0.99). Besides, it is also significantly upregulated in colon adenocarcinoma (FC=1.73), liver hepatocellular carcinoma (FC=1.68), cholangiocarcinoma (FC=1.63), esophageal carcinoma (FC=1.56), bladder urothelial carcinoma (FC=1.53), and lung squamous cell carcinoma (FC=1.50). CBX2 is associated with the proliferation of hepatocellular carcinoma by regulating the phosphorylation of YAP protein. Its overexpression is also linked with tumor progression in breast cancer due to its involvement in the activation of the PI3K/ATK signaling pathway. Therefore, overexpression of CBX2 can be targeted for therapeutic management of these cancers by developing effective and specific inhibitors of CBX2.

GLIOBLASTOMA GENE NETWORK RECONSTRUCTION AND ONTOLOGY ANALYSIS

Gubanova N.V.[1], Dergilev A.I.[2], Kononov V.A.[3], Orlova N.G.[4,5], Y.L. Orlov[3,6]

[1]Institute of Cytology and Genetics SB RAS, 630090 Novosibirsk, Russia
[2]Novosibirsk State University, 630090 Novosibirsk, Russia
[3]Agrarian and Technological Institute, Peoples' Friendship University of Russia, Moscow, Russia
[4]Financial University under the Government of the Russian Federation, Moscow, Russia
[5]Moscow State Technical University of Civil Aviation, Moscow, Russia
[6]I.M. Sechenov First Moscow State Medical University of the Ministry of Health of the Russian Federation (Sechenov University), Moscow, Russia

Glioblastoma is the most aggressive type of brain tumors resistant to a number of antitumor drugs. The problem of therapy and drug treatment course is complicated by extremely high heterogeneity in the benign cell populations, the random arrangement of tumor cells, and polymorphism of their nuclei. The pathogenesis of gliomas needs to be studied using modern cellular technologies, genome- and transcriptome-wide technologies of high-throughput sequencing, analysis of gene expression on microarrays, and methods of modern bioinformatics to find new therapy targets. Functional annotation of genes related to the disease could be retrieved based on genetic databases and cross-validated by integrating complementary experimental data. Gene network reconstruction for a set of genes (proteins) proved to be effective approach to study mechanisms underlying disease progression. We used online bioinformatics tools for annotation of gene list for glioma, reconstruction of gene network and comparative analysis of gene ontology categories. The available tools and the databases for glioblastoma gene analysis are discussed together with the recent progress in this field.

1. Dergilev, A.I., Orlova, N.G., Dobrovolskaya, O.B., Orlov, Y.L. (2021) J*ournal of Integrative Bioinformatics.* 18, 20200036.

PRELIMINARY *IN SILICO* STUDY OF 1,3-DIINDOLYLPROPENONES AND 1,3-DIARYLPROPENONES AS JNK ACTIVATORS

***J. Panada*[1], *V. Klopava*[1], *Y. Faletrov*[1,2], *T. Kulahava*[3], *N. Frolova*[1], *V. Shkumatov*[1,2]**

[1]*Research Institute for Physical Chemical Problems of the Belarusian State University, Minsk, Belarus*
[2]*Belarusian State University, Minsk, Belarus*
[3]*Institute for Nuclear Problems of the Belarusian State University, Minsk, Belarus*

Synthetic indoles are actively developed and studied due to their activity against cancer and a broad range of modes of antiproliferative action. Recently, substituted indoles and indolylchalcones were reported to induce methuosis, a form of non-apoptotic cell death mediated by JNK1/2 signaling and characterized by disruption of macropinocytosis and vacuolization of cell compartments, in U251 glioblastoma and several other cancer cell lines [1, 2]. In this work, we report the preparation of six 1,3-diindolylpropenones and 1,3-diarylpropenones by Claisen-Schmidt condensation (Fig. 1) and the docking calculations for phosphatidylinositol-3-phosphate 5-kinase (PIP5K3, PIKfyve), a regulator of JNK1/2 signaling and the proposed molecular target of methuosis-inducing indolylpropenones [3]. The results are shown in Table 1.

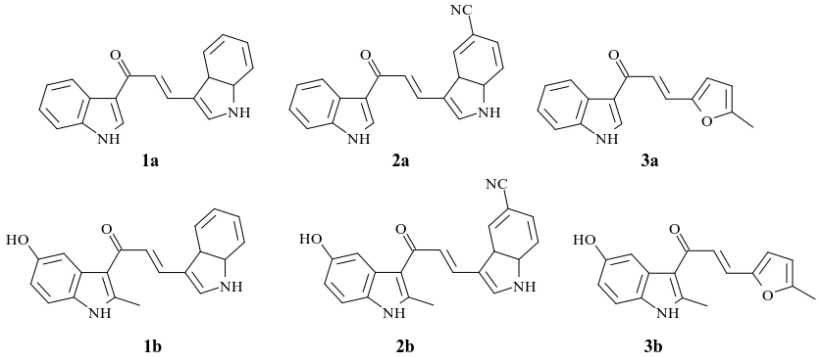


Figure 1. Structure of investigated 1,3-diarylpropenones

Table 1. Predicted interaction energies between PIP5K3 kinase and 1,3-diarylpropenones

Ligand	E_{bind}, kcal/mol	H-bonds	Ligand	E_{bind}, kcal/mol	H-bonds	Ligand	E_{bind}, kcal/mol	H-bonds
1a	-8.3[b]	-	**1b**	-8.9[b]	-	Apilimod	-8.7	O – Leu^{119}, NH – Asp^{215} (hydrazone)
2a	-9.3	5'-CN – Leu^{119}	**2b**	-8.7[a]	5-OH – Leu^{119}			
3a	-8.1[b]	-	**3b**	-7.7[a]	5-OH – Ala^{88}			

[a] Predicted to bind in the 3Å vicinity of the active site
[b] Predicted to localize within ≥ 5Å of the active site.

The best fit was predicted for **2a** and **2b**, with binding energies comparable to or above those for the reference PIP5K3 inhibitor apilimod. Notably, the positioning of **2a** within the active site allows a hydrogen bond between the 5'-cyano group and Leu^{119} amino group, which has also been described in literature for apilimod and could contribute to inhibitory activity. In summary, the results provide an insight into the structure required for effective binding to PIP5K3, which could be used for the rational design of new PIP5K3 inhibitors for glioma treatment.

This work was supported by Belarusian Republican Foundation for Fundamental Research grant X20M-113.

1. Huang, W., Sun, X., Li, Y. et al. (2018) *J. Med. Chem.*, 61, 5424-5434.
2. Cho, H.,Geno, E., Patoor, M. et al. (2018) *ACS Omega*, 3, 6097-6103.
3. Li, Z., Mbah, N. E., Overmeyer, J. H. et al. (2019) *BMC Cancer*, 19, 77.

WEB SERVICE FOR ASSESSMENT OF PUTATIVE METABOLITES AND PARENT COMPOUNDS USING MNA/QNA SIMILARITY

***A. Rudik*[1], *A. Dmitriev*[1], *A. Lagunin*[1,2], *D. Filimonov*[1], *V. Poroikov*[1]**

[1]*Institute of Biomedical Chemistry, Moscow, Russia*
[2]*Department of Bioinformatics, Pirogov Russian National Research Medical University, Moscow, Russia*

Since the metabolism of drug substance in the human organism often performed in several consecutive steps, almost every particular structure could be considered as both product of the previous biotransformation reaction and as a substrate of the next biotransformation reaction. Thus, solving of both direct and reverse problems in prediction of pharmaceutical substance metabolism is necessary in research and development of new medicines. We have developed a web service, which helps assess putative metabolites and parent compounds (http://way2drug.com/retro). Using the structure of a molecule entered by the user as a query, similar structures are identified in the known metabolic networks. As an output, the user obtains the list of similar compounds (Fig. 1, right part) with an opportunity to display the complete metabolic pathway (Fig. 1, central part).

The biotransformations knowledgebase includes over 2000 metabolic graphs compiled from DrugBank (https://go.drugbank.com), MetXBIODB [1] and ChEMBL (https://www.ebi.ac.uk/chembl) databases. We have used two methods to estimate the similarity. The first method uses MNA descriptors [2] and Tanimoto coefficient. The second method uses QNA descriptors [3] and the modified Tanimoto coefficient. Figure 1 shows an example for Diclofenac acyl glucuronide as a query. The resulting table contains the names and structures of parent compounds, structures of similar compounds and similarity estimates calculated by both methods.

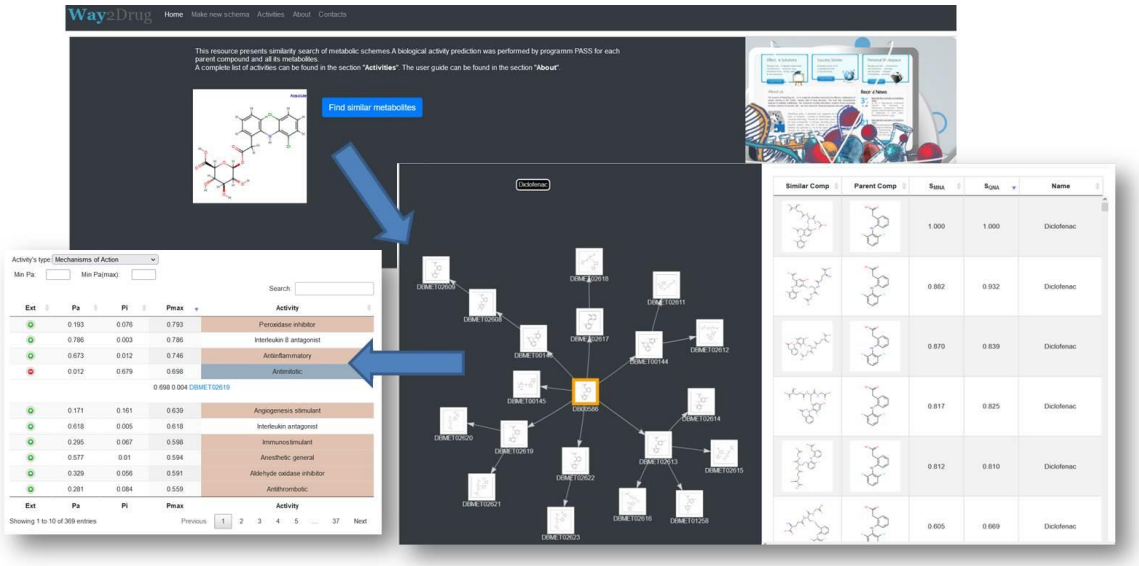


Figure 1. The example of the search for similar compounds for Diclofenac acyl glucuronide

In addition, for each compound from the known metabolic pathway, an integrated biological activity profile is calculated (Fig. 1, left part) [4].

This study was supported by the Russian Science Foundation grant No.19-15-00396.

1. Djoumbou-Feunang, Y., Fiamoncini, J., Gil-de-la-Fuente, A. et al. (2019) *J. Cheminform.*, 11, 2.
2. Poroikov, V.V., Filimonov, D.A., Gloriozova, T.A. (2019) *Russian Chemical Bulletin*, 68, 2143.
3. Filimonov, D.A., Zakharov, A.V., Lagunin, A.A. et al. (2009) *SAR & QSAR Environ. Res.*, 20, 679.
4. Rudik, A., Dmitriev, A., Lagunin, A. et al. (2021) *Molecular informatics*, 40, 2000231.

STRUCTURE BASED INHIBITOR DESIGN AGAINST SARS-COV-2 VIRUS USING THE ANALOGS OF NAFAMOSTAT AND CAMOSTAT THROUGH THE STUDIES OF DOCKING, MOLECULAR DYNAMICS AND DFT CALCULATION: AN IN-SILICO APPROACH

A. Muneeshwari, T. Sriram, N. Sampath

Chemical Biology Lab, School of Chemical and Biotechnology,
SASTRA Deemed to be University, Thanjavur, India

The severe acute respiratory syndrome-coronavirus-2 (SARS-CoV-2) is a viral infectious disease, which was first found in Wuhan city, China in December 2019. WHO declared in March 2020 that SARS-CoV-2 (also called as COVID-19) is a worldwide pandemic that spreads rapidly among the human. To control the viral infection, there are several vaccines have been developed and people were vaccinated, but no effective vaccines or drugs have been developed to date against this deadly virus. At the same time, SARS-CoV-2 has reemerged as several variants and spreading among the humans worldwide. Until today, approximately 500 million people have infected and 6 million people were dead across the world due to this lethal virus. Still, this infection rate is keep increasing across the world and creates panic around the people. However, this topic is still very hot among the pharmaceutical industries to develop efficient antiviral agent against COVID-19 virus. According to the bioinformatics analysis of SARS-CoV-2 spike protein with other viral structures of spike proteins, MERS-CoV and SARS-CoV showed high similarity among their RBD structures. As the RBD of spike protein is essential to bind with human Angiotensin Converting Enzyme 2 (ACE2) receptor, this event helps viral entry into the human host cell. Hence, this specific molecular recognition of virus with the host cell is a promising therapeutic drug target to discover a novel COVID-19 antiviral drug. Development of efficient vaccines or antiviral drugs against covid-19 may take longer period. Instead, repurposing of old FDA approved drugs could be the rapid and most effective treatment strategy to reduce the infection rate. Therefore, high throughput structure based virtual screening approach was utilized to examine the analogs of FDA approved Nafamostat and Camostat against viral spike protein (RBD), as they have identified as potential drugs against MERS-CoV. The top five drug leads from each were subjected to ADME analysis to evaluate the toxicity and efficacy of the drugs, DFT calculation for electronic structure properties and molecular dynamic (MD) simulation to understand the drug binding mode and stability at the active site of RBD (SARS-CoV-2), where the human receptor ACE2 binds. Therefore, further analysis of these drugs may insight into the effectiveness against the baleful virus.

MODELING, SYNTHESIS AND STUDY OF IMPACT OF N-t-BOC-(S)-β-[4-ALLYL-3-(PYRIDIN-3'-YL)-5-THIOXO-1,2,4-TRIAZOL-1-YL]-α-ALANYL-(S)-ALANYL-GLYCYL-(S)- β-[4-ALLYL-3-(PYRIDIN-3'-YL)-5-THIOXO-1,2,4-TRIAZOL-1-YL]-α-ALANINE TETRAPEPTIDE ON COLLAGENASE ACTIVITY

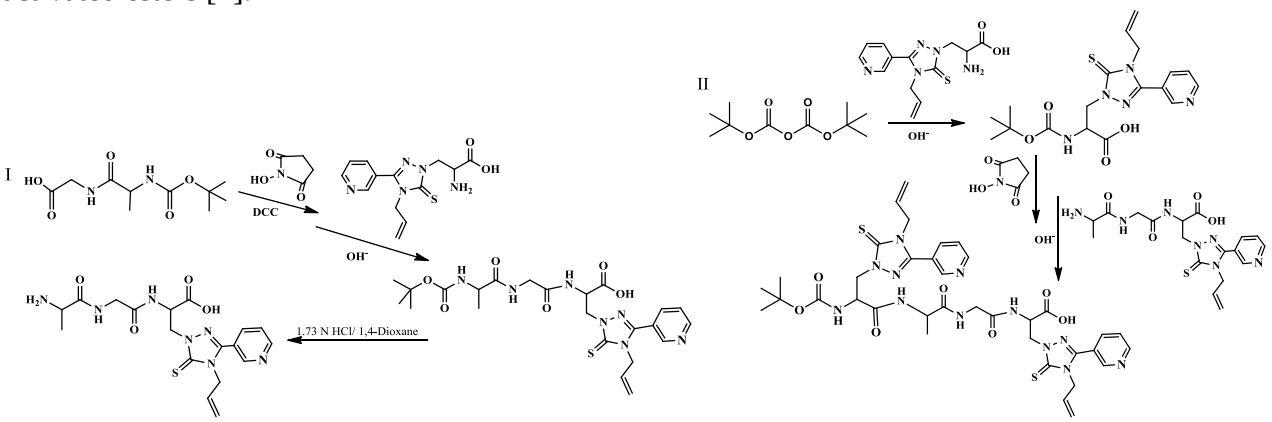
***A.S. Sargsyan*[1], *T.H. Sargsyan*[1,2], *Yu.M. Danghyan*[1], *S.M. Jamgaryan*[1], *E.A. Gyulumyan*[1], *A.M. Hovhannisyan*[2], *N.A. Hovhannisyan*[1,2], *A.H. Tsaturyan*[1,2], *A.S. Saghyan*[1,2]**

[1]*Scientific and Production Center "Armbiotevhnology" NAS RA, Yerevan, Armenia*
[2]*Yerevan State University, Yerevan, Armenia*

About 250 peptides with protective and free groups were constructed. Modeling of possible impact on collagenase activity of constructed peptides was carried out by molecular docking (AutoDock Vina) [1].

The most suitable molecule which is able to make complexes with enzyme is tetrapeptide- N-t-Boc-(S)-β-[4-allyl-3-(pyridin-3'-yl)-5-thioxo-1,2,4-triazol-1-yl]-α-Ala-(S)-Ala-Gly-(S)-β-[4-allyl-3-(pyridin-3'-yl)-5-thioxo-1,2,4-triazol-1-yl]-α-Ala with ΔG (Gibbs free energy) value –8.6 kcal/mol and KD (dissociation constant) value 0.497 μM.

Then, the synthesis of the above mentioned tetrapeptide was carried out by the method of activated esters [2].



The impact of synthesized peptides on collagenase activity was studied in vitro by the well-known method [3]. By using various peptide concentrations the value of IC_{50} was determined, which was 43 μM.

This study was supported by the Science Committee of RA, in the frames of the research project № 21T-2I235.

1. Trott, O., Olson, A. J. J.(2010) *Comput. Chem*., 31, 455–461.
2. Danghyan, Yu.M., Sargsyan, T.H., Djamgaryan, S.M., Gyulumyan, E.A., Panosyan, H.A., Saghiyan A.S. (2010) *Armenian Chemichal Journal*, 63 (1), 95-100. (in Rus.)
3. Mkrtchyan, A.F., Saghyan, A.S., Hayriyan, L.A., Sargsyan, A.S., Karapetyan, A.J., Tovmasyan, A. S., Tsaturyan A. H., Minasyan E. V., Poghosyan A. S., Paloyan A. M., Panosyan H. A., Sahakyan L.Yu. (2020) *Journal of Molecular Structure* 1208,. 127850.

DEVELOPMENT OF MACHINE LEARNING MODEL TO PREDICT CARDIOTOXICITY

***E.V. Shaniiazova*[1], *A.V. Golovin*[2]**

[1]*Lomonosov Moscow State University, Faculty of Medicine, Moscow, Russia*
[2]*Lomonosov Moscow State University, Faculty of Bioengineering and Bioinformatics, Moscow, Russia*

Research question. Development of 60% of new molecules closed due to cardiotoxicity. Moreover, the drugs Astemizole, Terfenadine, Cisapride (hERG inhibitors) were withdrawn, or their use was restricted. hERG (Human ether-à-go-go related gene) K^+ voltage-dependent channels ($K_v11.1$) blockage plays the critical role of interruption in cardiac action potential repolarization and leads to QT interval prolongation, Torsades de Pointes arrhythmia, and sudden death [1]. In vitro testing of hERG activity include different standard methods (PatchXpress, fluorescence polarization assay, etc.), but the experiments remain complex and expensive.

The aim of the present work is to design an in-silico machine learning classification model to predict cardiotoxicity and discuss the results.

Approaches and methods. The plan includes preprocessing of data, statistical and similarity-analysis, selecting descriptors, building a model based on classic machine learning algorithms, analysis of results.

Data. Data were collected from open ChEMBL database 29 v. with activities in IC_{50} to human hERG channel (13 200 rows). The potency threshold IC_{50} cutoff was set as 10 μM. Steps of preprocessing beside standard consist of excluding potency to the mutant form of hERG, and agonists activity. As a result, we got the balanced training dataset consisting of 5 952 compounds (48,2% blockers), the validation set of 1 489 compounds (45% blockers) and the test set with 296 well-known drugs.

Results. The designed model based on ensemble RandomForestClassifier and achieved the ROC_AUC=0.86 on test set. We used the descriptors ECFP3. The statistical analysis showed the peaks in the normal distribution of pIC_{50} related to values signed as threshold (10 μM and 30 μM). The classification model can bypass this issue. The similarity analysis allowed us to detect about 90 stereoisomers in training and test sets with the mean difference in IC_{50} about 22 μM (which is twice the value of threshold). Removing of stereoisomers with different activities from train set increased the quality of the model. The model correctly classified 74% of drugs in a test set: Astemizole, Dofetilide, Pimozide, Terfinadine, Cisapride, Verapamil, Haloperidol, etc. (hERG-blockers) and Nifedipine, Phenytoin, Nicotine, Phenobarbital, Prazosin, Loratadine, etc. (hERG-nonblockers). With using the feature importance with permutation method, we detected the most prominent topological descriptors: numbers of tertiary SP3 amines, LogP: octanol/water partition coefficient, TPSA (total polar surface area). The obtained results support the pharmacophore hypothesis of hERG-blocker: the presence of one tertiary charged nitrogen (forming cation-π interactions with residue Y652) and increased lipophilicity (provided by aromatic groups and forming hydrophobic contacts with residues Y652 and F656 [2]).

1. Konda, L.S.K., Praba, S.K., Kristam, R (2019) *Computational Toxicology*, 12, 100089.
2. Gualdani, R. et al. (2018) *Cellular Physiology and Biochemistry*, 45(6), 2233-2245.

SYNTHESIS AND IN SILICO EVALUATION OF BIOLOGICAL PROPERTIES OF A CIPROFLOXACIN-CHOLESTEROL CONJUGATE

L. Glinskaya[1], Y. Faletrov[1,2], V. Shkumatov[1,2]

[1]Research Institute for Physical Chemical Problems, Belarusian State University, Minsk, Belarus
[2]Faculty of Chemistry, Belarusian State University, Minsk, Belarus

Ciprofloxacin (CPF) is a fluoroquinolone antibiotic which is active against gram-positive and gram-negative bacteria and mycobacteria. From the other hand CPF ability to fluoresce with blue light and presence of fluorine atom together with both secondary amine and carboxyl moieties in its structure provide an option to use CPF as a label to detect it in complex environments using fluorescence-based and ^{19}F-NMR-based approaches [1-3]. Because of our interest to obtain a labeled cholesterol / oxysterol analogue aiming to gain an instrument to study or modulate biological functions of this class of molecules we performed synthetic conjugation of available CPF hydrochloride with cholesterol chloroformate which resulted in known compound CCF-CPF [4]; use of dimethyl sulfoxide/triethylamine mixture as solvent was a trick which allowed to mix together the both compounds with an extremely different polarity. Purity and identity of CCF-CPF product was confirmed using TLC (Rf=0.8 vs Rf=0.1 for CPF in SiO_2/acetone:acetic acid 4:1) and electrospray-mass spectrometry $[M+H]^+$ 744.50, $[M+Na]^+$ 766.50, $[M+triethylamine]^+$ 347.00).

Due to CCF-CPF was tested with respect to its anti-mycobacterial effect only [4], we have evaluated *in silico* its biological properties using PerMM server (for phospholipid membrane permeability) [5] and inverse high-throughput virtual screening using Autodock Vina [6] in pair with a helper tool FYTdock [7]. 1000 randomly chosen PDB structures of sterol-binding proteins were used. We obtained the following PerMM calculation results (parameters: pH 7.35, T=37°C): Free energy of binding was equal to -7.46 kcal/mol, log of permeability coefficients for plasma membrane, BBB and Caco-2 models were equal to -0.55, -3.09, and -3.62, respectively. These results prove quite good permeability of phospholipid membranes for CCF-CPF. Docking results revealed many hits and top-5 with docking scored from -15 to -14 kcal/mol includes PDB structures of cytochromes P450 CYP51 (PDB 5frb, 4zdy, 3jus) and lysosomal proteins NPC1 (6w5t) and LIMP2 (5uph).

This study was supported in the framework of GPSR (Belarus) No. 20210560.

1. Zimmermann, E.S., Torres, B.G., Dalla Costa, T. (2016) *Biomed Chromatogr.*, 30, 330.
2. Faletrov, Y.V., Glinskaya, L.I., Horetski, M.S. et al. (2021) *J. Belarusian State Univ. Chem.*, 1, 21.
4. Gee, C.T., Arntson, K.E., Urick, A.K. et al. (2016) *Nat Protocols*, 11, 1414.
4. Cilliers, P., Seldon, R., Smit, F.J. et al. (2019) *Chem Biol Drug Des*., 94, 1518
4. Faletrov, Y.V., Pozniak, H.I., Yakovets, P.S. et al. (2022) *Proceedings of the National Academy of Sciences of Belarus, Chemical Series*, 58, 62.
5. Lomize, A.L., Pogozheva, I.D., Mosberg, H.I. (2011) *J. Chem. Inf. Model.*, 51, 930.
6. Trott, O., Olson, A.J. (2010) *J. Comp. Chem.*, 31, 455.
7. Faletrov, Y.V., Staravoitava, V.A., Dudko, A.R. et al. (2022) PREPRINT (Version 1) available at Research Square [https://doi.org/10.21203/rs.3.rs-1456627/v1].

MOLECULAR DOCKING OF LIGANDS TO THE ACTIVE SITE OF DEOXYURIDINE TRIPHOSPHATASE

***A.G. Shults*[1], *V.R. Khairullina*[2], *Y.Z. Martynova*[1], *A.G. Mustafin*[1,2]**

[1]*Department of Chemistry, Bashkir State University, Ufa, Russia*
[2]*Ufa Institute of Chemistry of the Russian Academy of Sciences, Ufa, Russia*

Using molecular docking, the results of positioning 10 ligands with a common structural formula (Fig. 1) in the active site of deoxyuridine triphosphatase (dUTPase) were obtained. In the course of the studies performed using the evaluation function of the AutoDock 4.2.6 software package, the active optimal conformations for all ligands in the active site of dUTPase were determined and the numerical values of the efficiency of binding of these ligands to the active site of the enzyme were determined.

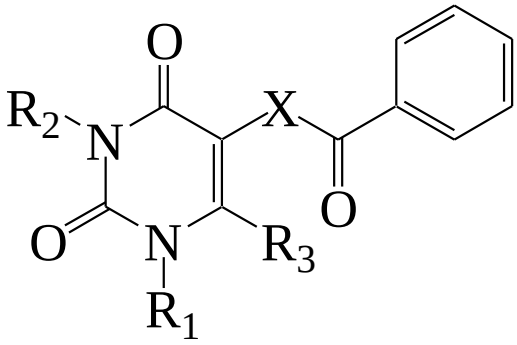


Figure1. General structural formula of ligands.

The preferred docking solutions were those ligand conformations that were in an energetically favorable state. As a result of molecular docking for 10 ligands, the thermodynamic characteristics of binding (E_{bind} and K_{inh}) to the dUTPase active site were evaluated. E_{bind} values lie in the range -6.29 - -4.24 kcal/mol, K_{inh} in the range 24.55-775.4 μmol.

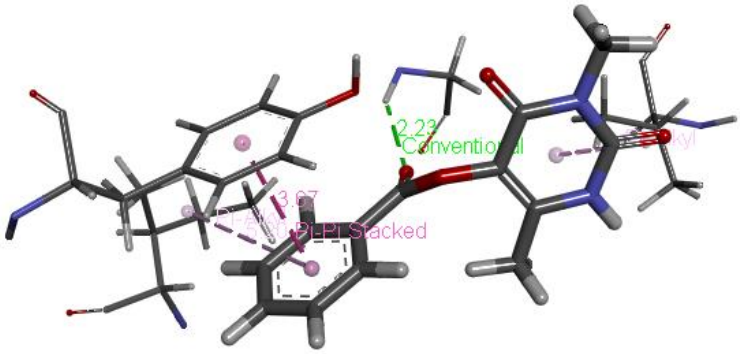


Figure 2. Interactions of ligands with amino acid residues

It was revealed that all ligands have hydrophobic interactions with the amino acid residue TYR105 (pi-pi-staking), as well as with two amino acids VAL65 and ILE101 (pi-Alkyl), (Fig. 2).

The reported study was funded by the Russian Science Foundation for Basic Research according to the research project No. 19-73-20073.

PHARMACOKINETICS PREDICTION OF LEADS FROM DIETARY SOURCES AGAINST NOVEL TARGETS OF PCOS AND PCOD

***A. Singh*, M.K. Selvam, A. Mukherjee**

Department of biotechnology, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, India

Conditions affecting several women across the globe, PCOS and PCOD are both endocrine disorders affecting ovary function. Even with multiple women grappling with PCOS/PCOD, minimal progress has been made concerning etiology and diagnosis [1-3]. Current management of symptoms associated with PCOS and PCOD involves lifestyle changes and prescribing contraceptive pills [2]. To ameliorate PCOS and PCOD at their roots, this report aims to identify key pathways and genes in PCOS/PCOD using an in-silico approach for screening lead therapeutic molecules to derive a novel treatment mechanism. For lead identification, compounds were taken from both FlavorDB and literature sources. Lead molecules were identified and optimized using ADMET properties. Genes linked to PCOS/PCOD-related anomalies (general abnormalities of the ovary) were determined using GeneCards and literature sources, both were used to obtain potential targets. Cytoscape was used to visualize the PPI network of the targets. The hub genes derived from the PPI network and the ones derived from the literature search were docked with the lead molecules. The study showed that the gene identified from the PPI network construction yielded the best docking score of -4.2 with lead number 2 from the list of ADMET screened molecules, among other ligand-target interactions. The gene mentioned above, derived from literature studies, is a histone methyltransferase and is a relevant target for this study owing to the connection established between epigenetic factors such as aberrant DNA methylation, non-coding RNAs, and the pathology of PCOS. Thus, this particular target has the most potential for bringing about the desired therapeutic effect pertaining to the prevention of random DNA methylation, thereby relieving PCOS/PCOD symptoms.

1. Deswal, R., Narwal, V., Dang, A., Pundir, C. S. (2020) *Journal of human reproductive sciences*, 13(4), 261–271.
2. Ganie, M. A., Vasudevan, V., Wani, I. A., Baba, M. S., Arif, T., Rashid, A. (2019), 150(4), 333–344.
3. PCOD and PCOS: Causes, Symptoms, Differences and Treatment. (n.d.). Pace Hospitals. URL: https://www.pacehospital.com/pcod-polycystic-ovary-disease-cause-symptoms-and-treatment

COMPUTATIONAL ANALYSIS OF KEY PLAYERS INVOLVED IN THE HEDGEHOG SIGNALLING PATHWAY

S. Ivanov[1], N. Biziukova[1], O. Tarasova[1]

[1]Institute of Biomedical Chemistry, Moscow, Russia

Malfunctioning of key proteins involved in the Hedgehog signaling pathway (Hh-pathway) has an essential role in the development of malignant neoplasms. There are many neoplastic diseases with activated Hh-pathway, including those that are characterized by a poor outcome [1].

The goal of our study is the search for human protein targets, including receptors, kinases, transcription factors, and some other possible master regulators of the Hedgehog signaling pathway. These protein targets may (1) have an impact on the proteins of the Hh-pathway; (2) transcription of genes, encoding these proteins; (3) genes, that can be up- or downregulated by the transcription factors GLI-1-3, that may be considered as the key elements of the Hh-pathway [2].

Using the originally developed approach for the text and data mining based on CRF and naïve Bayes approach [3, 4], we identified human protein and genes that are involved in the Hedgehog signaling pathway or can modulate its key targets. To reach this goal we automatically downloaded texts of publications abstracts relevant to the mechanisms of interactions between SHH, DHH, IHH, GLI-1, GLI-2, GLI-3 and any other human protein molecules; we have analyzed over 5000 publications. To identify proteins and genes that can be regulated by GLI-1, GLI-2, GLI-3, we considered only abstracts where names of proteins (genes) occurred along with GLI1-3. The interactions were detected using a set of rules developed based on a set of patterns (for instance, "regulated by", "activation of … by"), which help to consider interactions and its direction (for instance, phrases "protein1 activates protein 2" and "protein2 is activated by protein 1" are different is the meaning of interaction directions).

As a result, we identified 153 proteins (and corresponding genes names) interacting with GLI-1; 59 molecules interacting with GLI-2; and 47 proteins interacting with GLI-3. The proteins interacting with GLI-1 include cyclin-dependent kinase 7, mitogen-activated protein kinase 1, tyrosine 3-monooxygenase. The set of proteins interacting with GLI-2 includes B-cell lymphoma/leukemia 11B, Bcl2-associated agonist of cell death, cyclin-D1-binding protein 1. The proteins interacting with GLI-3 include ski oncogene, angiotensinogen, and nuclear receptor coactivator 4. For the majority of the proteins (genes) their role in tumor development [5, 6] or apoptosis regulation [7] has been shown in experimental studies.

This work was financed by the Ministry of Science and Higher Education of the Russian Federation within the framework of state support for the creation and development of World-Class Research Centers "Digital Biodesign and Personalized Healthcare" (No. 75-15-2022-305).

1. Fattahi, S., Nikbakhsh, N., Ranaei, M. et al. (2021) *Sci Rep*, 11, 7471.
2. Sigafoos, A., Paradise, B., Fernandez-Zapico, M. (2021) *Cancers (Basel)*, 13, 3410.
3. Biziukova, N., Tarasova, O., Ivanov, S., Poroikov, V. (2020) *Front. Genet*, 11, 618862.
4. Tarasova, O., Ivanov, S., Filimonov, D. A., Poroikov, V. (2020) *Molecules*, 25, 2944.
5. Permatasari, H., Nakahata, S., Ichikawa, T. et al. (2022) *Exp Hematol* S0301-472X(22)00158-8.
6. Niwa, Y., Kamimura, K., Ogawa, K. et al. (2022) *J Clin Med*, 11, 851.
7. Boonyarat, C., Sangchavee, K., Plekratoke, K. et al. (2021) *Biol Pharm Bull*, 44, 494.

A NOVEL IMPROVED APPROACH FOR IN SILICO DESIGN OF EFFECTIVE AND NONTOXIC SARS-COV-2 INHIBITORS

***R.S. Tumskiy*[1], *A.V. Tumskaia*[2], *I.N. Klochkova*[2]**

[1] *Institute of Biochemistry and Physiology of Plants and Microorganisms – Subdivision of the Federal State Budgetary Research Institution Saratov Federal Scientific Centre of the Russian Academy of Sciences, Saratov, Russia*

[2]*Chemistry Institute, Saratov State University, Saratov, Russia*

The global pandemic caused by SARS-CoV-2 is one of the challenging outbreaks the world has ever experienced. Despite on known mutations of this virus, the main protease (Mpro) and the papain-like protease (PLpro) are promising specific viral targets, which highly conserved for SARS-CoV-2. Importantly note that partial inhibition of several targets can be more effective in comparison with full inhibition of a single target [1]. For this reason, the present work is aimed on discovery of new dual inhibitors of Mpro and PLpro using a novel improved approach for virtual design.

In the previous studies a prospective inhibitor of Mpro (azachalcone derivative **94**) was identified *in silico* [2]. Despite on high affinity to viral protease by docking data, this compound has unfavorable ADMET properties (in particular, binder **94** potentially acts as thiol-reactive compound). For these reasons, in this work azachalcone **94** is selected as a starting structure for *in silico* design of dual inhibitors of Mpro/PLpro with potentially favorable ADMET profiles.

For the first time, we propose a novel improved multistep approach for *in silico* design of effective and nontoxic inhibitors based on combination of next methods: an artificial neural network-driven platform LigDream (a shape decoding tool), ADMET prediction, consensus docking and molecular dynamics. In the first step of *in silico* design, LigDream generated 100 new molecules from starting structure **94**. In the second step, we selected only one potentially nontoxic compound (pyrazolopyridazine derivative **89**) with good pharmacokinetics according to ADMET prediction. In the third step, LigDream was used for generation derivatives of compound **89** (16 pyrazolopyridazines with excellent ADMET profiles). In the next step, consensus docking (GOLD with Autodock Vina) of pyrazolopyridazines was performed within active site of Mpro/PLpro proteins. The docking poses for one of the discovered inhibitor with active sites of SARS-CoV-2 proteins are depicted in Fig. 1.

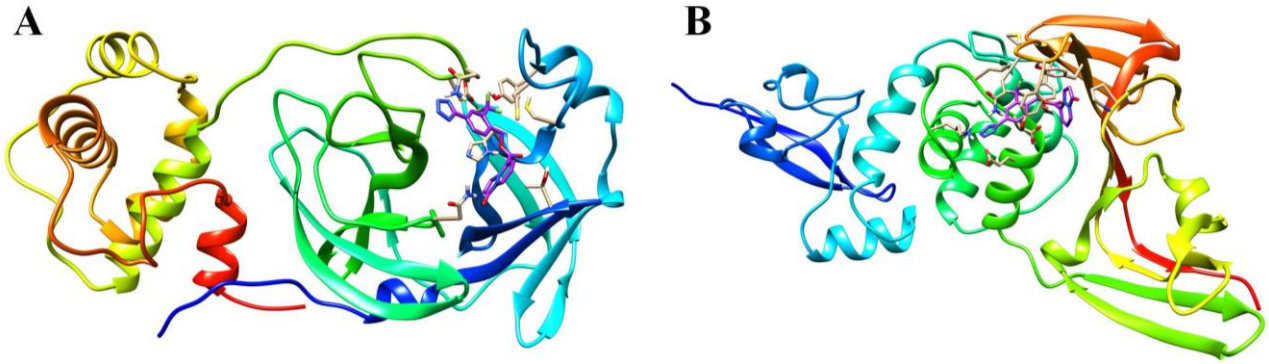


Figure 1. (A) The docking pose of inhibitor with Mpro target; (B) The docking pose of inhibitor with PLpro target

Finally, molecular dynamics simulation was used to evaluate stability of protein-ligand complexes. As a result of multistep design *in silico*, five novel nontoxic inhibitors with potentially high affinity to Mpro/PLpro proteins were identified.

Our provided results can be used for discovery of new antiviral drugs for the treatment of COVID-19.

1. Csermely, P., Agoston, V., Pongor, S. (2005) *Trends in pharmacological sciences*, 26(4), 178-182.
2. Tumskiy, R.S., Tumskaia, A.V. (2021) *Chemical Physics Letters*, 780, 138894.

AFFINITY PREDICTION OF PROTEIN COMPLEXES FROM STRUCTURE-BASED FEATURES USING GRADIENT BOOSTING

***D.I. Tychinin*, K.V. Shaitan, V.N. Novoseletsky**

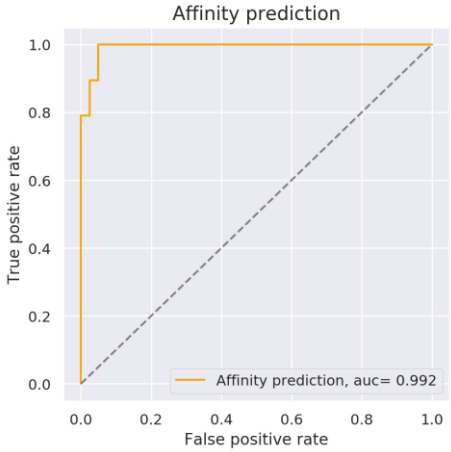
Moscow State University, Faculty of Biology, Moscow, Russia

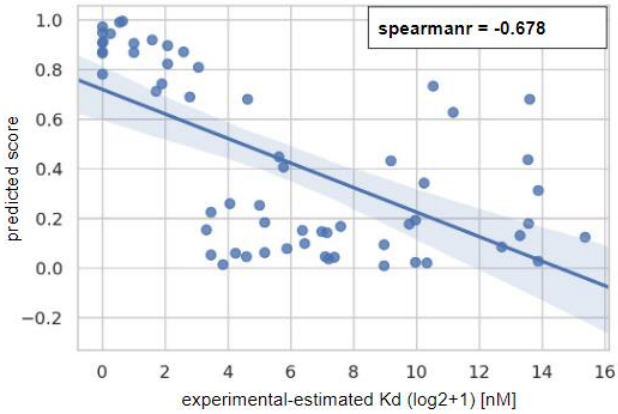
The problem of computational methods for the affinity prediction of protein-protein interactions remains relevant. We have developed a new model for computational estimation affinity based on gradient boosting using the LGBM software package [1]. Model training was performed on 591 crystallographic structures of protein complexes selected from PDBbind [2]. Selection criteria:

1) the resolution of the crystallographic structures must be less than 3 angstroms;
2) Debye-Waller factor for atoms of the crystal structure should not exceed 80 along the median.

The complexes are divided into affinity classes according to experimentally known interaction constants. Computationally, the protein-protein interaction in each complex was represented as a matrix of atomic group contacts, according to the typing of heavy atoms in the CHARMM force field [3]. Next, the pairwise distances between the two protein structures of the complex were calculated. Distances were calculated using the MDAnalysis software package [4].

The set of test data was 10% of the entire sample (60 complexes). The AUC of the resulting model was 0.992. Test data metrics in the table below. The obtained predictive score also inversely correlates with the experimentally measured Kd (spearman coefficient = - 0.678, p-value < 0.05). The resulting model works on various protein targets, as well as in a wide range of interaction constants.





Class	Precision	Recall	F1-score	n-Samples
Kd >7 nM	1.00	0.88	0.94	41
Kd <7 nM	0.79	1.00	0.88	19

1. Guolin, Ke, et.al. (2017) *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 3149–3157
2. URL: pdbbind.org.cn
3. Vanommeslaeghe, K. et. al (2010) *J. Comput. Chem.*, 31(4), 671–690.
4. Richard, J., Gowers et al. (2016) *Proceedings of the 15th Python in Science Conference*, 98–105.

IDENTIFICATION OF POTENTIAL NEW DRUGGABLE TARGET IN *SALMONELLA TYPHI* BIOFILM

***A. Upadhayay*[1], *D. Pal*[2], *A. Kumar*[1]**

[1]*Department of Biotechnology, National Institute of Technology, Raipur, India*
[2]*Department of Chemical Engineering, National Institute of Technology, Raipur, India*

Biofilm of *Salmonella Typhi* confers the capacity for colonization and long-term persistence of infection. It is considered as one of the most pressing public health concerns and rising of antibiotic resistance is resulted into high mortality. Most powerful antibiotic medicines are failed to control this infection due to Biofilm formation. *Salmonella Typhi* infection has also been linked with other lethal diseases like gall bladder cancer. As a result, the *S. Typhi* outbreak is becoming an incurable illness. The interplay of several genes and proteins are responsible for this pathogen's virulence. Therefore, this study was aimed to identify some essential drug target in *S. Typhi* so that new drug could be designed against this pathogen to address the problem. In this study, we have identified 15 putative target proteins that played an essential role in *S. Typhi* Biofilm development and maturation. Three proteins CsgD, AdrA, and BcsA were shown to have a role in the formation of cellulose, a crucial component of the extracellular matrix of Biofilm. Result further indicated that the CsgD protein had significant interconnectivity and strong interaction with other essential target proteins of *S. Typhi*. As a result, it is deciphered and concluded that CsgD plays function in a variety of processes, including cellulose formation, bacterial pathogenicity, bacterial quorum sensing, and bacterial virulence. The hydrophobic property and cellular localization of all identified target proteins were also explored and results additionally provided the evidence of prospective therapeutic target based on computational and bioinformatics study. Result is overcoming the concept of drug target-shortage in *S. Typhi*, and we deduce the possibility of their future design and development as promising anti-salmonella therapeutics.

ELECTROPHILIC FRAGMENT SCREENING FOR NEW PROTEIN LABELS AND POTENTIAL NONPEPTIDIC INHIBITORS OF CATEPSINS

***P. Yakovets*[1,2], *Y. Faletrov*[1,2], *V. Shkumatov*[1,2]**

[1]*Research Institute for Physical Chemical Problems, Belarusian State University, Minsk, Belarus*
[2]*Faculty of Chemistry, Belarusian State University, Minsk, Belarus*

Cysteine is a unique proteinogenic amino acid with a strongly nucleophilic thiol group. Because of relatively rare abundance in proteins and participation in various catalytic processes, protein modification by cysteine residue is widely used aiming to such proteins structure and functions studies [1], mimicking of posttranslational modifications [2], proteomics and covalent inhibitor design [3]. It is known that compounds with -CH_2-Br fragment are able to react with thiols quickly and selectively. We synthesized few such compounds (Figure 1).

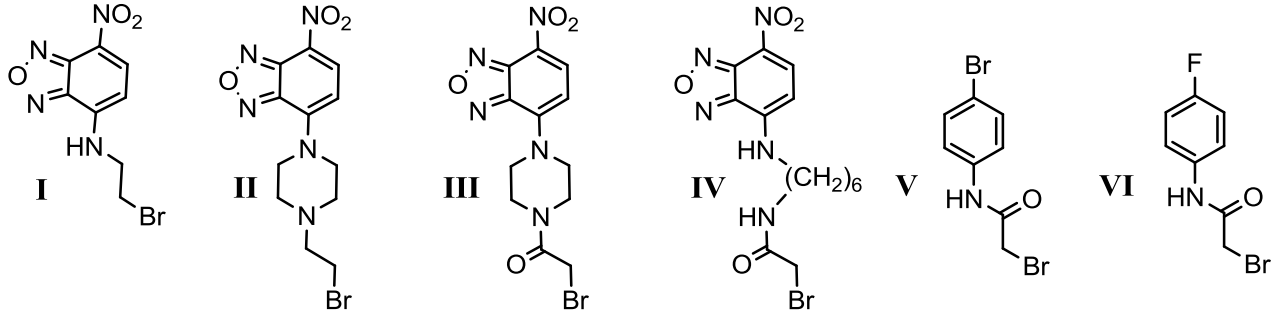


Figure 1. Structures of the compounds under consideration

Structures of the compounds are partially confirmed using mass-spectrometry and ^{1}H-NMR. Compounds I-IV possess 7-nitrobenzoxadiazol-4-amine group and, thus, protein modification by the compounds will result in fluorescent conjugates [4], whereas compounds V and IV possess bioortogonal halogen atoms and, thus, their conjugates with proteins could have advantages for analysis using high resolution mass-spectrometry, ^{19}F и ^{79}Br/^{81}Br NMR [5], X-ray studies [6]. An additional analysis of potential of the compounds in the frame of fragment-based inhibitor discovery approach we conducted docking of them with few PDB structures of cathepsins – lysosomal cysteine proteases being considered as drug targets. Using Autodock Vina together with helper tool FYTdock we have revealed affine binding of II with cathepsin S (PDB codes 3n4c, 2h7j, 2op3) with energy of binding (docking score) values of from -7.7 to -7.3 and colocalization of the compound's bromine atom close to S-atom of CYS25. Because of known ability of NBD-piperazine (a fragment of II-IV) to be taken up by lysosomes, the predicted ability of II provide perspectives for experimental evaluation.

This study was supported by GSPSR (Belarus) No. 20210560.

1. Akabas M.H. (2015) *Adv Exp Med Biol*., 869, 25.
2. Simon, M. D., Chu, F., Racki, L. R. et al. (2007) *Cell*, 128, 5.
3. Wu S., Howard H.L., Wang H. et al. (2016) *Biochem Biophys Res Commun*, 478, 1268.
4. Faletrov, Y.V., Karpushenkova, V.S., Zavalinich, V.A., Yakovets, P.S. (2021) *J. Belarusian State Univ. Chemistry*, 2, 25.
5. Gee C.T., Arntson K.E., Urick A.K. et al. (2016) *Nat Protocols*, 11, 1414.
6. Zimmermann M.O., Lange A., Wilcken R. et al. (2014) *Future Med Chem*. 6, 617.
7. Faletrov, Y.V., Staravoitava, V.A., Dudko A.R. et al. (2022) PREPRINT (Version 1) available at Research Square [https://doi.org/10.21203/rs.3.rs-1456627/v1].

CREATION OF SAR MODELS FOR PREDICTION OF PATHOGENIC AMINO ACID SUBSTITUTIONS IN PROTEINS RELATED TO MONOGENIC DISEASES INCLUDED IN NEWBORN SCREENING

***A. Zadorozhnyy*[1], *A. Smirnov*[1], *D. Filimonov*[2], *A. Lagunin*[1,2]**

[1]*Department of Bioinformatics, Pirogov Russian National Research Medical University, Moscow, Russia*

[2]*Department of Bioinformatics, Institute of Biomedical Chemistry, Moscow, Russia*

Over the past few years, next-generation sequencing (NGS) technologies have been actively implemented into clinical practice. In particular, there are successful results using NGS in newborn screening (NBS) [1]. The largest NBS genetic program takes place in United States, with approximately 4 million infants screened annually [2]. The American College of Obstetricians and Gynecologists' (ACOG) Committee on Genetics Conditions published "Recommended Uniform Screening Panel of Core Conditions", which mainly includes diseases caused by the malfunction of specific protein. Due to the screening wild coverage and the availability of NGS, it is highly likely that new genetic variants without interpretation will be discovered. The most common variants requiring clinical classification are missense, when one amino acid (a.a.) is replaced by another. This work is intended to present the structure–activity relationship (SAR) analysis of a.a. substitutions and their surroundings in specific proteins to predict clinical effect of the variants as additional interpretation. From 32 core conditions from ACOG screening panel 23 monogenic diseases were chosen and the associated genes were found based on OMIM database. Related to the genes known missense variants annotated data, including clinical significance, variant supporting evidence and protein allele were obtained using BioMart data mining tool [3]. Variants currently classified as pathogenic or likely pathogenic constituted a positive class, and substitutions that were interpretated as benign/likely benign, as well as all those that were in no way related to the phenotype/disease, constituted the negative class. Datasets contain fix length peptides (7-31 a.a. in a peptide) from the substitution and a.a. surroundings in the form of structural formulas in MOL V3000 format plus their effect indicators (0-benign, 1-pathogenic). The similar algorithm was early used for prediction of phosphorylation sites in proteins [4]. Amino acid surroundings were taken from canonical reference protein' sequences by related positions. Classification models were created and validated in a modified version of Prediction of Activity Spectra for Substances (OnLineMultiPASS) software [4,5], which allows using different levels of multilevel neighborhoods of atoms (MNA) descriptors to describe the peptide's chemical structure. Each of the fifteen MNA levels was used to build the individual model on each of thirteen different peptide fragment lengths dataset.

According to values of invariant accuracy of prediction (IAP) we chose SAR models with optimal levels of MNA descriptors and peptide length. In case of close scores, we chose classifiers built on lower parameters. As a result, SAR models with acceptable IAP values from 0.70 to 0.96 calculated by 20-fold cross-validation for nineteen diseases were selected (e.g. hemoglobinopathies, biotinidase deficiency, classic phenylketonuria and some others).

This work was supported by grant No. 075-15-2019-1789 from the Ministry of Science and Higher Education of the Russian Federation

1. Remec, Z.I., Trebusak Podkrajsek, K., Repic Lampret, B. et al. (2021) *Front. Genet.* 12, 662254.
2. Newborn screening and the role of the obstetrician–gynecologist. ACOG Committee Opinion No. 778. (2019) American College of Obstetricians and Gynecologists.
3. Kinsella, R.J., Kähäri, A., Haider, S. et al. (2011) *Database (Oxford),* 2011.
4. Karasev, D.A., Savosina, P.I., Sobolev, B.N. et al. (2017) *Biomed Khim.*, 63(5), 423-427.
5. Filimonov, D.A., Lagunin, A.A., Gloriozova, T.A. et al. (2014) *Chem Heterocycl Comp,* 50, 444–457.

7-NITROBENZOXADIAZOL-4-YL *ORTHO*-PHENYLENEDIAMINE DERIVATIVE AS A NEW MULTITARGET LIGAND FOR METALLOPROTEINS: *IN SILICO* EVALUATION

Y. Faletrov[1,2], V. Zavalinich[1], M. Gvozdev[1,2], N. Loginova[1,2], V. Shkumatov[1,2]

[1]Research Institute for Physical Chemical Problems, Belarusian State University, Minsk, Belarus
[2]Faculty of Chemistry, Belarusian State University, Minsk, Belarus

Ortho-phenylenediamine (OPD) has redox-properties allowing him to be substrates for peroxidases [1] and, like o-catechols and o-aminophenols, for cytochromes c [2]. Both cytochromes c and peroxidases have hems with iron ion as redox-active groups. Also, N-acyl-OPD derivatives are among histone deacetylases' (HDACs) inhibitors, including drugs. HDACs have Zn^{2+} complexed with side chains of His and Asp residues of the enzymes [3]. From the other hand, 7-nitrobenzoxadiazole derivatives of amines are able to absorb visible light, fluoresce and they are interesting as ligands for various proteins (e.g., see [4]); N-NBD-para-aminophenol was published as a peroxidase substrate [5].

We have synthesized an OPD derivative bearing one 7-nitrobenzoxadiazol-4-yl group (NBD-OPD hydrochloride) using commercially-available NBD-Cl and OPD. Its purity was confirmed using TLC, spectrophotometry. To evaluate biological properties of NBD-OPT we conducted Autodock Vina docking-based inverse high-throughput virtual screening using FYTdock to organize, run and analyse results of the screening. Sets of PDB structures of cytochromes c and Zn-containing beta-lactamases, totally 500 structures (Table 1). Energy of binding values (Ebind) were found to be from -10.5 (for RNase, PDB 3bk2) to -5.3 (for cytochrome c, PDB 3o1y). For metallo-beta-lactamases (PDB 2hb9, 5lcf, 5mm9 and 4ua4; Ebind= -9.1 - -8.8) the free $-NH_2$ group in calculated complexes were close to Zn of the enzymes. Cytochrome c structure PDB 6cun was found to demonstrate close localization of the -NH2 group and Fe. Further docking search for other potential target proteins. The received data provide initial roadmap for a way to choose for experimental researches.

This study was supported by GSPSR (Belarus) No. 20210560.

1. Hu W.-C., Pang J., Biswas S. et. al. (2021) *Anal. Chem.*, 93, 8544.
2. Loginova N.V., Koval'chuka T.V., Faletrov Y.V. (2011) *Polyhedron,* 30, 2581.
3. Lauffer B.E. L., Mintzer R., Fong R. et al. (2013) *J. Biol. Chem.* 288, 26926.
4. Faletrov, Y.V., Karpushenkova, V.S., Zavalinich, V.A. et al. (2021) *J. Belarusian State Univ. Chemistry*, 2, 25.
5. Wang C. et al. (2013) *Analyst*., 10, 2825.

CONJUGATES OF PYROPHEOPHORBIDE a WITH STEROIDAL SCAFFOLD. SYNTHESIS, MOLECULAR MODELING, INTERACTION WITH SOME CANCER CELLS

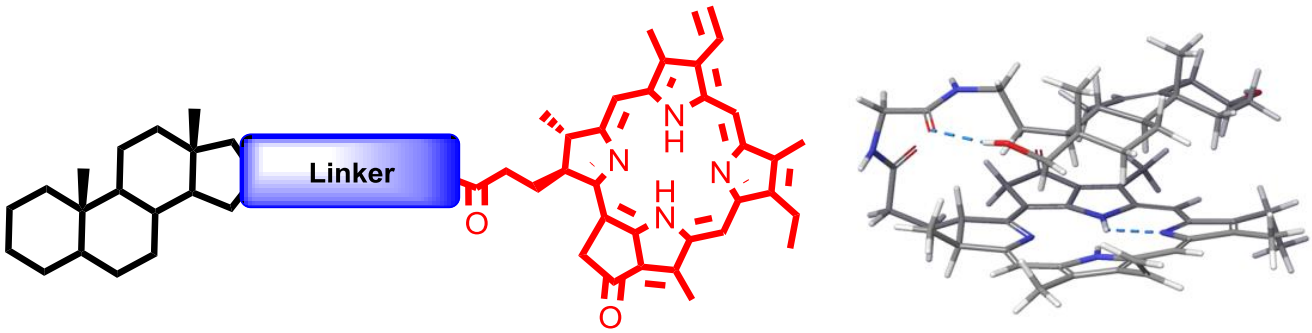
A. Korolchuk[1,2], A. Lukin[1], G. Morozevich[1], R. Novikov[3], Y. Tkachev[3], V. Zolottsev[1,4]

[1]Institute of Biomedical Chemistry, Moscow, Russia;
[2]Institute of Fine Chemical Technologies, Moscow, Russia;
[3]Engelhardt Institute of Molecular Biology, Moscow, Russia;
[4]RUDN University, Moscow, Russia

Tetrapyrrolic macrocycles porphyrins and chlorins owing to their unique photochemical and photophysical properties have wide range of biomedical applications such as optical imaging, fluorescent labeling, photodynamic inactivation of microbial infections, and photodynamic therapy of solid tumors. Coupling of macrocycles with drugs or fragments of biological active molecules significantly improves delivery and distribution of macrocycle-based compounds to a specific location within the cells, facilitates their transport through receptor or drug mediated endocytosis, increased their specificity and selectivity, affects their photo chemical properties and biological activity.

Conjugation of pheophorbide a or pyropheophorbide a with steroids is considered to be a promising approach for development of new bifunctional constructs possessing enhanced delivery to specific targets.

In this study we have synthesized and investigated seventeen new conjugates of pyropheophorbide a with androgen receptor ligands – testosterone, dihydrotestosterone, epitestosterone, 3β-hydroxy-seco-D-androstene and seco-D-testosterone.



Spectral properties and molecular models of conjugates revealed a significant influence of the structure on the conformation of conjugates. The study of interaction of conjugates with prostate carcinoma LNCaP and PC-3 cells indicated that all conjugates were efficiently uptaken and internalized by cells and potently inhibited their growth and proliferation. Anti-proliferative activity of same conjugates, as well as their photo induced toxicity in prostate carcinoma cells was dependent on the length of linker and the stereochemical configuration of steroidal part.

We suggest that some of newly synthesized conjugates may be considered as prospective agents for various biomedical experiments in cultured cells.

This work was financed by the Ministry of Science and Higher Education of the Russian Federation within the framework of state support for the creation and development of World-Class Research Centers "Digital Biodesign and Personalized Healthcare" (No. 75-15-2022-305).

STRUCTURE-BASED VIRTUAL SCREENING OF ANTHOCYANINS WITH INHIBITORY POTENTIALS ON DIFFERENT MOLECULAR TARGETS OF DIABETES

***P.A. Akinnusi*[1], *S.O. Oluboda*[3], *A.A. Alade*[1], *A.J. Akinnusi*[1], *S.O. Olubode*[1], *O.Y. Olawumi*[2], *R.U. Adekola*[3], *A.A. Ashimi*[1], *S.A. Shodehinde*[1]**

[1]*Department of Biochemistry, Adekunle Ajasin University, Akungba-Akoko, Nigeria*
[2]*Department of Microbiology, Adekunle Ajasin University, Akungba-Akoko,Nigeria*
[3]*Department of forestry and wildlife, University of Benin, Benin City, Nigeria*

Type II diabetes is known to be a 21st-century epidemic and is responsible for approximately 90% - 95% of diabetes cases. The pathophysiological distortions are majorly β-cell dysfunction, insulin resistance, and chronic inflammation, which all progressively unsettle the control of blood glucose levels and trigger micro-and macrovascular complications. The wide range of pathological disruptions present in type II diabetes mellitus gives rise to the consideration that multiple antidiabetic agents, administered in combination, will be required to maintain normal blood glucose. Herein, a library of 118 anthocyanins was screened to identify small molecular weight compounds with inhibitory effects on Protein tyrosine phosphatase 1B (PTP1B), Dipeptidyl peptidase-4 (DPP-4), and α-amylase. From the result, the top five compounds with the highest binding affinity were reported. ADMET profiling showed moderate pharmacokinetic profiles for these compounds as well as insignificant toxicity. Cyanidin 3-(p-coumaroyl)-diglucoside-5-glucoside (-15.272 kcal/mol), Cyanidin 3-O-(6"-malonyl3"-glucosyl-glucoside) (-9.691kcal/mol), and Delphinidin 3,5-O-diglucoside (- 12.36 kcal/mol) had the highest binding affinities to PTP1B, DPP-4 and α-amylase respectively and can be used in combination to control glucose fluctuations. However, validations must be carried out.

MOLECULAR DOCKING AND VIRTUAL SCREENING REVEALS THE ROLE OF GLAFENINE HYDROCHLORIDE AS A NOVEL BIOMOLECULE FROM *BACILLUS PARALICHENIFORMIS* POSSESSING NEMATICIDAL PROPERTY AGAINST BANANA BURROWING NEMATODE *RADHOPHOLUS SIMILIS*

S. Ashraf[1], S. Nakkeeran[2], N. Saranya[3], R. Saravanan[2], V. Ragapriya[3], R.U. Krishna Nayana[1]

[1]*Department of Plant Biotechnology, Tamil Nadu Agricultural University, Coimbatore, India*
[2]*Department of Plant Pathology, Tamil Nadu Agricultural University, Coimbatore, India*
[3]*Department of Plant Molecular Biology and Bioinformatics, Centre for Plant Molecular Biology and Biotechnology, Tamil Nadu Agricultural University, Coimbatore, India*

Bananas are the developing world's fourth most important food crop (after rice, wheat and maize) in terms of gross value of production. However, its production is limited by numerous diseases and pests, such as various species of plant-parasitic nematodes, which infest banana roots and have negative repercussions (Mendoza and Sikora, 2009). Recently, there is a trend to search for biorational pesticides to overcome the hazardous effects of synthetic compounds necessitising the identification of an effective and eco-friendly biomolecule for the management of nematodes. Molecular modelling and docking studies were performed to discover an effective nematicidal biomolecule for the management of *Radophalus similis.* Glafenine hydrochloride is one of the secondary metabolite produced by *Bacillus paralicheniformis* during the ditrophic interaction with *Fusarium oxysporum* f.sp. *cubense* (*Foc*). In the present work, we have analyzed the nematidical property of glafenine hydrochloride against *Radopholus similis* protein targets calreticulin, Cathepsin S-like cysteine proteinase, β- 1,4 -endoglucanase and reticulocalbin, venom allergen-like protein, cytochrome c oxidase subumit 1 and serine carboxypeptidase. Three-dimensional modelled structure of protein targets of *R. similis* were docked with biomolecules through AutoDock Vina module in PyRx 0.8 server to predict the binding energy of ligand and target protein. Among the chosen six targets, docking analysis revealed that glafenine hydrochloride had the maximum binding affinity of -8.8 kcal/mol for reticulocalbin, -7.7 kcal/mol for calreticulin, -7.8 kcal/mol for β- 1,4 -endoglucanase , -9.0 kcal/mol for Cytochrome c oxidase subumit 1 , and -7.5 kcal/mol for serine carboxypeptidase in comparison with nematicide carbofuran 3G. Besides, increased binding affinity of glafenine hydrochloride with the protein target sites and thus have facilitated to explore it as a novel nematicidal molecule for the management of banana burrowing nematode *R. similis*.Thus, present investigation confirmed that, the small molecule glafenine hydrochloride can be explored for nematicidal activity.

OBSERVATION ON TWO-DIMENSIONAL APTAMERS FOLDING IN-SILICO

***M.Z. Sabri*[1], *C.A. Norbert*[1], *T.T. Chai*[2]**

[1]*Universiti Kuala Lumpur Malaysian Institute of Chemical and Bioengineering Technology, Melaka, Malaysia*

[2]*Universiti Tunku Abdul Rahman, Perak, Malaysia*

Aptamers are short single-stranded oligonucleotides and peptides that hold a specific binding affinity to a broad-spectrum target molecule such as ions, complex proteins, and antigens up to the cellular surface and cell organelles[1]. Aptamers are mainly extracted through restrictive in-vitro screening procedures known as Systematic Evolution of Ligands by EXponential Enrichment (SELEX). There is a great deal of interest and growing demand in basic and clinical sciences to identify, analyze, and quantify aptamers' tiny molecules and proteins. Nevertheless, the costs and time required to select aptamers demand a great reagent consumption and investment of time. This research aims to evaluate the performance of three computational tools in predicting the secondary structure of aptamer. The web-server-based programs of CentroidFold, RNAfold, and Mfold, were employed in this study. A total of 24 aptamers data were collected from the Protein Data Bank (PDB). The structure folding analysis was performed on the selected aptamers' 2-dimensional (2D) and 3-dimensional (3D) original structure and nucleotides arrangement. Subsequently, the Tanimoto similarity score was used to calculate the accuracy of programs to predict each aptamer's folding. The results showed that for 4 of the 24 aptamers, 100% matched between the PDB derived and web-server predicted structure was discovered using at least one of the algorithms, accounting for 17% of the dataset. The Mfold gave the most accurate prediction, with a Tanimoto score of 0.9. In conclusion, Mfold, RNAfold, and CentroidFold modeling tools are capable of predicting the secondary structure of aptamers.

This study was supported in the initiative of UniKL-UTP Collaboration Research Fund (CRGS Title: Development of DNA-based aptamer for effective detection of Hepatitis B surface antigen using electrochemical sensor).

1. Chen, Z., Hu, L., Zhang, B.T. et al. (2021) *International Journal of Molecular Science*, 22, 7.

AUTHOR INDEX OF ABSTRACTS

ISBN 978-5-900760-18-6