



BBCC2016

Bioinformatica e Biologia Computazionale in Campania

off the record
Abstract Book

Consiglio Nazionale delle Ricerche
Istituto di Scienze dell'Alimentazione

Avellino

16 Dicembre 2016

Responsabile del Convegno:

Angelo Facchiano

CNR-ISA – Consiglio Nazionale delle Ricerche, Istituto di Scienze dell’Alimentazione, Avellino

Comitato Scientifico:

- Angelo Facchiano – CNR-ISA, Avellino
- Claudia Angelini – CNR- IAC, Napoli
- Maria Luisa Chiusano – Università di Napoli “Federico II”
- Italia De Feis – CNR- IAC, Napoli
- Paola Festa – Università di Napoli “Federico II”
- Mario Guarracino – CNR- ICAR, Napoli
- Anna Marabotti – Università di Salerno
- Roberto Tagliaferri – Università di Salerno

Comitato Organizzatore:

Angelo Facchiano – CNR-ISA, Avellino

Anna Marabotti – Università di Salerno

Eugenio Del Prete – CNR-ISA, Avellino

Serena Dotolo – CNR-ISA, Avellino

Deborah Giordano – CNR-ISA, Avellino

Bernardina Scafuri – CNR-ISA, Avellino

Supporto tecnico:

Clemente Meccariello - CNR-ISA, Avellino

Antonio Ottombrino – CNR-ISA, Avellino

BBCC2016 è organizzato con il supporto di:

Progetto Bandiera InterOmics



con il patrocinio di BITS – Bioinformatics Italian Society



BBCC2016 – 16 Dicembre 2016, CNR-ISA, Avellino

Programma

10.00 Apertura registrazione e affissione poster ai pannelli

Prima Sessione

10.30 Introduzione al Convegno
Angelo Facchiano

10.50 Structural determinants of the residue conformational preferences: the key role of the peptide geometry
Nicole Balasco, Luciana Esposito, Alfonso De Simone, Luigi Vitagliano

11.10 CONSRANK, a pure consensus approach for the scoring of protein-protein docking models: results in latest CAPRI rounds and perspectives
Edrisse Chermak, Ida Autiero, Luigi Cavallo, Romina Oliva

11.30 Computational analysis and biological characterization of a newly synthesized cephalosporin analogue
Anna Verdino, Giovanni Vigliotta, Deborah Giordano, Ivana Caputo, Annunziata Soriente, Margherita De Rosa, Anna Marabotti

11.50 Understanding of the activity of a protein involved in DNA repair by biochemical, structural and in silico approaches
Castrese Morrone, Riccardo Miggiano, Mario Serpe, Alberto Massarotti, Anna Valenti, Giovanni del Monaco, Mosè Rossi, Franca Rossi, Menico Rizzi, Maria Ciaramella, Giuseppe Perugino

12.10 Molecular simulations for investigating the role of food polyphenols in human health
Bernardina Scafuri, Anna Marabotti, Angelo Facchiano

12.30 Bridging nanotoxicology and nanopharmacology: A systems biology approach
Dario Greco

13.00-14.20 Pausa pranzo e discussione poster

Seconda Sessione

- 14.20 Overview of Elixir Teaching activities 2017
Vincenza Colonna
- 14.40 Combining ChIP-Seq and Nascent-Seq to Investigate the
Oncosuppressor Activities of Estrogen Receptor Beta in Breast Cancer
Giorgio Giurato
- 15.00 Bioinformatic Approaches to Cancer Metabolomic Profiling
Analysis@CROM.INT-Pascale
*Angela Sorice, Francesca Capone, Eliana Guerriero, Gennaro Ciliberto,
Alfredo Budillon, Susan Costantini*
- 15.20 DecontaMiner: a tool for the identification of contaminating sequences in
human NGS sequencing data
Ilaria Granata, Mara Sangiovanni and Mario R Guarracino
- 15.40 Discovering novel and rare causative variants in undiagnosed diseases
using VarGenius, a software for variants detection and annotation
***Francesco Musacchia, Margherita Mutarelli, Andrea Ciolfi, Marco
Tartaglia, Sandro Banfi, Vincenzo Nigro***
- 16.00 Network-based dimension reduction methods for the
integrative analysis of multi-omics data in cancer
***Antonella Iuliano, Annalisa Occhipinti, Claudia Angelini, Italia De
Feis, Pietro Liò***
- 16.20 The solution of large-scale Minimum Cost SAT Problem as a tool for
data analysis in bioinformatics
*Giovanni Felici, Daniele Ferone, Paola Festa, Antonio Napoletano,
Tommaso Pastore*
- 16.40 Discussione e Conclusioni

Presentazioni orali e Posters

Structural determinants of the residue conformational preferences: the key role of the peptide geometry

Nicole Balasco (1), Luciana Esposito (1), Alfonso De Simone (2), Luigi Vitagliano (1)

(1) Institute of Biostructures and Bioimaging, C.N.R., Naples I-80134, Italy

(2) Division of Molecular Biosciences, Imperial College South Kensington Campus, London SW7 2AZ, UK

The definition of the structural basis of the conformational preferences of the genetically encoded amino acid residues is crucial to decipher the physical code of protein folding and would have a huge impact on our understanding of protein structure and function. Indeed, although a large number of computational and experimental investigations have highlighted that the different protein residues show distinct conformational propensities, none of the current hypotheses is able to satisfactorily explain these preferences.

In the last decades, we and others have clearly demonstrated that several geometrical parameters of protein backbone (bond angles, peptide bond distortions from planarity, and pyramidalization of the carbonyl carbon atom) are heavily dependent on the local conformation (ϕ ; ψ ; dihedral angles) [1-8]. Moreover, a correlation between bond distances such as CO and CN has been detected in ultrahigh resolution protein structures [9]. Concerning bond angles, most of these investigations have been focused on the $NC\alpha$; $C\tau$; angle, shown to be significantly affected by both ϕ ; ψ dihedral angles. In this framework, we here evaluated the impact of the local geometry on the residues conformational preferences by performing statistical analyses on a dataset of non-redundant protein chains selected from the Protein Data Bank (PDB). Our data clearly indicate that the peptide geometry has a key role in dictating residues propensities for the secondary structure elements. The results of these analyses will be illustrated.

1. Karplus PA. *Protein Sci.* 1996; 5:1406-20.
2. Berkholtz DS, Shapovalov MV, Dunbrack RL Jr, Karplus PA. *Structure*, 2009;17:1316-25.
3. Esposito L, Balasco N, De Simone A, Berisio R, Vitagliano L. *Biomed Res Int*, 2013;2013:326914.
4. Improta, R., L. Vitagliano, and L. Esposito, *Proteins*, 2015; 83:1973-86.
5. Improta, R., L. Vitagliano, and L. Esposito, *Acta crystallographica D*, 2015; 71:1272-83.
6. Esposito L, De Simone A, Zagari A, Vitagliano L. *J Mol Biol*, 2005; 347:483-7.
7. Berkholtz DS, Driggers CM, Shapovalov MV, Dunbrack RL Jr, Karplus PA. *Proc Natl Acad Sci*, 2012;109:449-53.
8. Esposito L, Vitagliano L, Zagari A, Mazzarella L. *Protein Sci*, 2000; 9:2038-42.
9. Esposito L, Vitagliano L, Zagari A, Mazzarella L. *Protein Eng*, 2000; 13:825-8.

CONSRANK, a pure consensus approach for the scoring of protein-protein docking models: results in latest CAPRI rounds and perspectives

Edrisse Chermak (1), Ida Autiero (1), Luigi Cavallo (1) and Romina Oliva (2)

(1) Kaust Catalysis Center, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia.

(2) Department of Sciences and Technologies, University “Parthenope” of Naples, Napoli.

Algorithms for scoring protein-protein docking decoys generally rely on physics-based and/or knowledge-based terms. We introduced in the field the first pure consensus method, CONSRANK (CONSENSUS RANKING) [1]. CONSRANK, also available as a web server [2], ranks models based on their ability to match the most conserved (or frequent) inter-residue contacts in the ensemble they belong to.

Blind testing in recent CAPRI Rounds, including the CASP11-CAPRI30 joint experiment, showed CONSRANK to perform on par with the state-of-the-art energy- and knowledge-based scoring functions, at least for targets with well-defined interaction interfaces and docking ensembles sufficiently enriched in correct solutions [3]. More recently, we implemented a modified approach, Clust-CONSRANK, embedding a contact-based clustering step prior to the CONSRANK algorithm. Clust-CONSRANK was developed in an attempt to overcome one of the intrinsic limitations of a pure consensus approach, i.e. the propensity to select similar solutions. Testing on past CAPRI targets has shown the clustering step to substantially enhance the chance of identifying at least one correct solution for the most challenging targets [4].

A combined CONSRANK/Clust-CONSRANK approach has been therefore applied to the scoring of protein-protein docking models in the latest CASP-CAPRI joint experiment, CASP12-CAPRI37, run during summer 2016. Preliminary assessment results show that, with our algorithms, we could identify native-like solutions for 8 assessed interfaces (corresponding to 6 targets), on par with the best performing scorers in this experiment. Moreover, we achieved the highest number of correct solutions overall (72 vs. 64 of the second scorer in such terms).

CONSRANK is available as a web server at:
<https://www.molnac.unisa.it/BioTools/consrank/>

Contact: romina.oliva@uniparthenope.it

1. Oliva, R., Vangone, A. and Cavallo, L. (2013) Ranking multiple docking solutions based on the conservation of inter-residue contacts. *Proteins* 81, 1571-1584.
2. Chermak, E., Petta, A., Serra, L., Vangone, A., Scarano, V., Cavallo, L., Oliva, R. (2015) CONSRANK: a server for the analysis, comparison and ranking of docking models based on inter-residue contacts, *Bioinformatics* 31, 1481-3.
3. Lensink, M.F. et al. (2016) Prediction of homo- and hetero-protein complexes by protein docking and template-based modeling: a CASP-CAPRI experiment, *Proteins* 84, Suppl 1:323-48.

4. Chermak, E., De Donato, R., Lensink M.F., Petta, A., Serra, L., Scarano, V., Cavallo L., Oliva R. (2016) Introducing a Clustering Step in a Consensus Approach for the Scoring of Protein-Protein Docking Models, PLoS One 11:e0166460.

Computational analysis and biological characterization of a newly synthesized cephalosporin analogue

Anna Verdino (1), Giovanni Vigliotta (1), Deborah Giordano (1,2), Ivana Caputo (1), Annunziata Soriente (1), Margherita De Rosa (1) and Anna Marabotti (1)

(1) Dept. Chemistry and Biology "A.Zambelli", University of Salerno

(2) Institute of Food Science, Italian National Research Council, Via Roma 64, 83100 Avellino and Ph.D. School of health food innovation and management, University of Foggia, Italy.

Antibiotics are among the most important drugs discovered in the 20th century, but their efficacy during the last decades has been seriously challenged by the development of bacterial strains resistant to their action. For this reason, there is the compelling need to develop new antimicrobial molecules with innovative solutions in order to bypass the bacterial defenses. In this work, we present the prototype of a new class of beta lactam antibiotics, in which the 7-aminocephalosporanic nucleus is functionalized with an isolated azetidinone ring. This new compound has been proved effective against Gram positive microorganisms, with a potency comparable or even better with respect to ceftriaxone, a third-generation cephalosporin widely used in clinics. Moreover, it has showed no cytotoxicity against two different human cell lines, thereby proving its safety even at a concentration 10 times higher than the minimum inhibitory concentration (MIC) found for (tested) Gram positive bacteria.

A deep computational analysis performed with an innovative covalent docking approach has been conducted with the aim of understanding the molecular effects and the contribution of the different moieties of this compound to the binding energy towards selected penicillin-binding proteins (PBPs) from both Gram positive and Gram negative organisms, including an enzyme typical of a multidrug resistant *Staphylococcus aureus* strain.

From the results, it appears that the compound is able to bind to all PBPs with similar predicted affinity, but the cephalosporin ring has a predicted binding energy lower than the one of the isolated azetidinone ring, thus it can be inferred that the former is the preferred moiety for the interaction with the enzyme. However, the substituents of the isolated azetidinone ring interact with the residues of the active sites, therefore their functionalization could in principle modulate the affinity of the compound. The two diastereoisomers obtained from the chemical synthesis, instead, seems not to be discriminated by the enzymes.

All these results will help us developing derivatives of this compound with improved chemical and biological properties, such as a broader spectrum of action and/or an increased affinity towards their molecular targets.

This work has been made in the frame of the projects FARB-ORSA 151138 and 161582 (for A. V., D. G., I. C., A. M.), FARB-ORSA 143020 and 158155 (for G.V.), FARB 2013 (for A. S. and M. D. R.).

Understanding of the activity of a protein involved in DNA repair by biochemical, structural and *in silico* approaches

Castrese Morrone (1), Riccardo Miggiano (2), Mario Serpe (1), Alberto Massarotti (2), Anna Valenti (1), Giovanni del Monaco (1), Mosè Rossi (1), Franca Rossi (2), Menico Rizzi (2), Maria Ciaramella (1), and Giuseppe Perugino (1)

(1) Istituto di Bioscienze e BioRisorse - UOS Napoli, CNR. Via P. Castellino 111, 80125 Napoli

(2) DiSF-Dip. di Scienze del Farmaco, Università del Piemonte Orientale, Via Bovio 6, 28100 Novara
giuseppe.perugino@ibbr.cnr.it

The repair of DNA from alkylation damage is generally performed by evolutionary conserved protein complexes. However, specific repair of O^6 -alkylated-guanines is a task of a small class of proteins called AGTs (alkylated DNA-protein alkyltransferases): by using a single-step reaction mechanism, the alkylic group is irreversibly transferred to a catalytic cysteine in the active site, inducing the *in vitro* and *in vivo* inactivation and destabilization of the protein. Although some conformational changes after the alkylation are supposed, a complete picture of structural rearrangements occurring during the reaction cycle is missing. The complete knowledge of these structural movements is a great challenge and a fundamental task for the development of new inhibitors of the human AGT, whose overexpression leads to a resistance in several types of tumour cells to the chemotherapeutic alkylating agents-based treatment.

We used the *Sulfolobus solfataricus* thermostable ortholog (*SsOGT*) as a model for AGTs [1], by performing biochemical, structural, molecular dynamics and *in silico* analysis of ligand-free, DNA-bound and alkylated version of the protein. With this protein, we were able to highlight conformational changes and perturbations of intramolecular interaction occurring during lesion recognition and catalysis, confirming our previous hypothesis that coordination between the N- and C-terminal domains of *SsOGT* is important for protein activity and stability [2]. All the data allowed us to propose a general model of structural rearrangements occurring during the reaction cycle of AGTs [3], and proposing it as a starting point to design strategies to modulate AGT activity in therapeutic settings.

References

[1] G. Perugino, A. Vettone, G. Illiano, A. Valenti, M.C. Ferrara, M. Rossi, M. Ciaramella (2012) Activity and regulation of archaeal DNA alkyltransferase: conserved protein involved in repair of DNA alkylation damage. *J. Biol. Chem.*, 287, 4222-4231.

[2] G. Perugino, R. Miggiano, M. Serpe, A. Vettone, A. Valenti, S. Lahiri, F. Rossi, M. Rossi, M. Rizzi, M. Ciaramella (2015) Structure-function relationships governing activity and stability of a DNA alkylation damage repair thermostable protein. *Nucleic Acids Res.*, 43, 8801-8816.

[3] C. Morrone, R. Miggiano, M. Serpe, A. Massarotti, A. Valenti, G. del Monaco, M. Rossi, F. Rossi, M. Rizzi, G. Perugino, M. Ciaramella (2017) Interdomain interactions rearrangements control the reaction steps of a thermostable DNA alkyltransferase. *BBA-Gen. Sub.*, 1861, 2, 86-96.

Molecular simulations for investigating the role of food polyphenols in human health

Bernardina Scafuri (1), Anna Marabotti (2), Angelo Facchiano (1)

(1) National Research Council, Institute of Food Science, Avellino, Italy

(2) Dept. Chemistry and Biology "A.Zambelli", University of Salerno, Italy

Food components are object of many studies aimed to identify compounds responsible of positive effects on human health. Compound categories include antioxidants, whose activity against pathologies as cancer and cardiovascular diseases is commonly related to their activity in deactivating free radicals. Polyphenols are of particular interest, and studies demonstrate many positive effects for human health [1].

We have recently investigated the role of apple polyphenols in chemoprevention of colorectal cancer [2], an effect reported in literature in the past years. We evidenced possible protein targets for polyphenols found in apple but present also in other fruits and vegetables.

Our results open the perspective of identifying metabolic pathways affected by polyphenols, with both positive and negative effects on human health. To this aim, we are investigating in more detail the potential protein targets already identified in our study, and defining the possible role of each compound investigated in the modulation of the related target activity. This ongoing study is based on molecular simulations of protein-ligand docking, mutation effects, protein-protein interactions, and pathways analysis.

References

1. Vauzour, D., Rodriguez-Mateos, A., Corona, G., Oruna-Concha, M. J. & Spencer, J. P. Polyphenols and human health: prevention of disease and mechanisms of action. *Nutrients*, 2010, **2**, 1106-1131.
2. Scafuri B, Marabotti A, Carbone V, Minasi P, Dotolo S, Facchiano A. A theoretical study on predicted protein targets of apple polyphenols and possible mechanisms of chemoprevention in colorectal cancer. *Sci Rep*. 2016 Sep 2;6, 32516. doi: 10.1038/srep32516.

Combining ChIP-Seq and Nascent-Seq to Investigate the Oncosuppressor Activities of Estrogen Receptor Beta in Breast Cancer

Giurato G (1,2)

1) Laboratory of Molecular Medicine and Genomics, Department of Medicine, Surgery and Dentistry 'Scuola Medica Salernitana', University of Salerno, Via S. Allende 1, 84081 Baronissi (SA), Italy;

2) Genomix4Life, University of Salerno, Via S. Allende 1, 84081 Baronissi (SA), Italy;

The Estrogen Receptor beta (ER β) is a member of the nuclear receptor superfamily of transcriptional regulators and it is involved in oncosuppressive activities, antagonizing hormone-induced carcinogenesis and inhibiting growth and oncogenic functions in luminal-like breast cancers (BCs), where its expression correlates with a better prognosis of the disease. Unliganded ER β is expressed in both nucleus and cytoplasm of BC cells and several evidences suggest that it can exert an active role in gene regulation at transcriptional and post-transcriptional level even in the absence of the hormone. Starting from these observations, interaction proteomics by Tandem Affinity Purification (TAP) coupled to mass spectrometry (MS) was applied in BC cells to identify cellular factors acting in concert with ligand-free ER β , leading to isolation of protein complexes associated with unliganded ER β . Among the proteins identified, our attention was focused on AGO2, due to the implication of this argonaute protein in several key biological processes [1].

To study the functional association of the complex ER β -AGO2, we performed ChIP-Seq analyses in MCF-7 cells expressing tagged ER β , to map common and unique binding sites of the two proteins to the genome. Subsequently, Nascent-Seq was performed under the same conditions before and after AGO2 silencing, to assess the potential role of this factor in affecting ER β -modulated transcription rate and RNA splicing.

Combination of ChIP-Seq and Nascent-Seq data revealed significant changes in the transcription rate and maturation of genes targeted together by both factors and involved in cell death and survival networks.

Reference:

[1] Tarallo R, Giurato G et al., 2016 (under revision)

Bioinformatic Approaches to Cancer Metabolomic Profiling Analysis@CROM.INT-Pascale

Angela Sorice (1), Francesca Capone (1), Eliana Guerriero (1), Gennaro Ciliberto (2), Alfredo Budillon (1), Susan Costantini (1)

(1) CROM, Istituto Nazionale Tumori “Fondazione G. Pascale” - IRCCS, Napoli, Italia

(2) Direttore Scientifico, Istituto Nazionale Tumori “Regina Elena”- IRCCS, Roma, Italia

Malignant cells undergo significant changes in metabolism including a redistribution of metabolic networks and resulting in different metabolic landscapes in cancer cells versus normal cells. Metabolomics is a global approach useful to identify the overall metabolic changes associated with a particular biological process, to find the most affected metabolic networks, and to provide an additional layer of information that can be linked with transcriptomics and proteomics data.

In the last years cancer metabolomics studies were performed using metabolic fingerprinting and profiling by NMR spectroscopy on biological fluids (sera and urine), tissue extracts or cell pellets. Using NMR spectroscopy techniques being not invasive and not expensive, it is possible to differentiate several tumor types, to identify metabolites specific for cancer initiation and progression, and to study the drug effects on cancer patients in order to select useful prognostic markers.

Metabolomics generates large complex and multi-dimensional amounts of data in similar manner to other “omics” approaches. Hence, storing, pre-processing and multivariate statistical analysis of these data provide a significant challenge and require specialized bioinformatics tools. In particular, data and associated metadata require to develop specific databases whereas for multivariate statistical analysis of the metabolomics data it is possible often to use the same approaches applied to the analysis of other genomic and proteomic data. However, metabolomics needs also of different tools in respect to others common “omics” data due to the fact that they are generated by multiple analytical platforms and require extensive data pre-processing. In general, further bioinformatic developments in metabolomics data analysis include raw analytical data processing (bucketing), metabolomics standards and ontology, statistical analysis and data mining, pathway analysis, and development of classifier model by mathematical modeling of metabolic networks within the framework of systems biology.

Our group at CROM – INT Pascale is evaluating the metabolomics profiling by ¹H-NMR and applying the related bioinformatics tools: i) to study the effects of drug treatment on human cancer cell lines for identifying the metabolic pathways modulated from these molecules, and ii) to search on cancer patient sera the metabolites decreasing and increasing during the cancer progression.

DecontaMiner: a tool for the identification of contaminating sequences in human NGS sequencing data

Ilaria Granata, Mara Sangiovanni and Mario R Guarracino

ICAR-CNR, Via Pietro Castellino 111, 80131, Napoli, Italy

Reads mapping on a reference genome is a crucial step of NGS (Next Generation Sequencing) data analyses. The reads that fail to align are usually discarded and considered as not informative.

Characterizing the unmapped reads and understanding their source is essential to assess the quality of the whole experiment and to get some insights on possible “contamination” from non-human sequences (e.g. viruses, bacteria, fungi).

Downstream contamination may occur during the samples processing, whereas upstream contamination might be due to the presence of microorganisms infecting the sampled tissues. Here we present DecontaMiner, a tool for the detection of viral, bacterial and fungi contamination in human sequenced data. A local alignment strategy is used to detect similarities between input reads and putative contaminating organism sequences.

Several outputs are generated to provide a thorough and exhaustive report of the unaligned reads.

DecontaMiner has successfully detected foreign sequences in human cancer RNA-seq data.

Discovering novel and rare causative variants in undiagnosed diseases using VarGenius, a software for variants detection and annotation

Musacchia F (1), Mutarelli M (1), Ciolfi A(2), Tartaglia M(3), Banfi S (1), Nigro V (1)

(1) Telethon Institute of Genetics and Medicine (TIGEM) via Campi Flegrei, 34 - 80078 Pozzuoli (NA)

(2) Centro di Ricerca per gli Alimenti e la Nutrizione CREA Via Ardeatina 546, 00178 Roma (Italy)

(3) Genetics and Rare Diseases Research Division Ospedale Pediatrico Bambino Gesù

VarGenius is a pipeline designed to allow fast and accurate variant discovery from exome and targeted sequencing data. It has been designed for the Telethon Undiagnosed Disease Project (UDP) whose aim is to find an explanation for diseases which are still mysterious by using Whole Exome Sequencing (WES). The software is written in PERL language and uses the UDP protocol for the analysis: BWA is used for alignment against reference genome [1], GATK for variant discovery [2] and Annovar for the annotation of variants [3].

VarGenius takes in input a text file which contains information about the samples: kinship, path of the fastq files, sample name and analysis name plus few other parameters useful during the execution and a configuration file containing all the parameters used for the analyses. Data about the samples and the variants information are stored in a PostgreSQL database which is useful for computing allele frequencies and for the organization of the analyses. Genotypes can be calculated using entire families or the joint calling technique which exploits many different samples to obtain highly accurate variant calling. The software is optimized for fast execution with PBS cluster by executing a different job for each sample in each analysis.

Since the project has started, 24 trios (or quartets) from the UD project have been sequenced with Illumina technology and aligned against the HG19 human reference to detect novel and rare causative variants. We prioritized causative variants in both known and unknown disease genes that are currently being validated.

The possibility to analyze dozens of samples in parallel using a single command, the efficiency of usage of cluster resources in terms of memory and threads allocation and finally the possibility to create a database with variants information, makes this software unique among the variant discovery and annotation tools.

Further development of the pipeline are oriented towards the usage of the newest HG38 human reference and extending the analysis to whole genome sequencing data.

[1] Li H. and Durbin R. (2009) Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics*, 25:1754-60. [PMID: 19451168]

[2] Mark A DePristo, Eric Banks, Ryan Poplin, Kiran V Garimella, Jared R Maguire, ChristopherHartl, Anthony A Philippakis, Guillermo del Angel, Manuel A Rivas, Matt Hanna, Aaron McKenna, Tim Fennell, Andrew M Kernysky, Andrey Y Sivachenko, Kristian Cibulskis, Stacey B Gabriel, DavidAltshuler & Mark J Daly. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics* 43,491–498 doi:10.1038/ng.806

[3] Wang K, Li M, Hakonarson H. ANNOVAR: Functional annotation of genetic variants from next-generation sequencing data *Nucleic Acids Research*, 38:e164, 2010

Network-based dimension reduction methods for the integrative analysis of multi-omics data in cancer

Antonella Iuliano (1), Annalisa Occhipinti (2), Claudia Angelini (1), Italia De Feis (1), Pietro Liò (2)

(1) Istituto per le Applicazioni del Calcolo "Mauro Picone", Napoli, Italy

(2) Computer Laboratory, University of Cambridge, UK

Cancer is one of the most complex diseases at multiple molecular levels. No single level of genomic data can fully explain tumor behaviour, necessitating analysis at multiple levels encompassing genomics and proteomics. In the last years, thanks to international projects and consortia, the access to the genome-wide data at multiple molecular levels has been made available by a variety of high-throughput technologies. For example, The Cancer Genome Atlas and the European Genome-phenome Archive were established to profile large tumor sets at both DNA and RNA levels and to create integrated repositories of all types of sequences and genotype experiments, including the aberrations present in cancer cells. Hence, these data can provide opportunities and challenges to develop statistical and computational tools for the analysis, interpretation and validation of cancer data and help the research community to improve the prevention, diagnosis, and treatment of cancer.

However, from a statistical perspective, the most important challenge integrating multi-omic analyses is the high-dimensionality of the data. In fact, taking more levels into account increases the dimensionality of the problem. In particular, adding more layers of data raises the dimension of unknown parameters, which are often difficult to estimate, thereby making the overall inference weaker. In addition, at every step, there are checkpoints of data compatibility, such as normalisation at the same scale, adequate correction for technical batch effects and use of different platforms. As a consequence, to tackle such problems, we propose a new multistage computational-statistical strategy for the integration of omics data into survival regression models. We fuse dimension reduction procedures and network-penalized Cox models for high dimensional survival data. First, we integrate different data types identifying the best approach for combining multiple matrices that include data from different scales and then, we reduce the high-dimensionality of data. Second, we incorporate gene regulatory network information using penalized Cox regression methods. Third, we test the predictive power of the selected gene signatures using independent datasets to derive novel disease gene interaction pathways and disease-risk genes. By using this approach, it is possible to obtain a deeper insight of the gene-regulatory networks and investigate the gene signatures related to the survival time of cancer patients.

The solution of large-scale Minimum Cost SAT Problem as a tool for data analysis in bioinformatics

G. Felici (1), D. Ferone (2), P. Festa (2), A. Napoletano (2), and T. Pastore (2)

(1) CNR, Institute for Systems Analysis and Computer Science, IASI-CNR 00185 Rome, 6 Italy

(2) University of Napoli “Federico II”, Dept. of Mathematics and Applications, 80126 – Napoli, Italy

Data mining is one of the main activities in bioinformatics, specifically to extract knowledge from massive data sets related with gene expression measurement, CNV, DNA strings, and others.

A long array of methods are used to perform such task, ranging from the more established parametric statistical analysis to non-parametric techniques, to classification methods that have been developed in knowledge engineering and artificial intelligence.

In this paper, we consider a method for extracting logic formulas from data that relies on a large body of literature in integer and logic optimization, originally presented in [1], that has been largely and successfully applied to different problems in bioinformatics ([2], [3], [4], [5], [6]).

Such method is based on the iterative solution of Minimum Cost SAT Problems and is able to extract logic formulas in DNF form that possess interesting features for their interpretation. While leaving the discussion of the main features and motivations of this approach to the related literature, in this talk we focus on the problem of solving efficiently very large scale instances of this well-known logic programming problem and propose a new GRASP approach that, being able to exploit the specific structure of the problem, largely outperforms other established solvers for the same problem.

References

1. G. Felici, K. Truemper. A Minsat Approach for Learning in Logic Domains, *INFORMS Journal on Computing* 14(1): 20-36 (2002).
2. P. Bertolazzi, G. Felici, E. Weitschek. Learning to classify species with barcodes, *BMC Bioinformatics*, 10:1-12 (2009).
3. M. Arisi, R. D’Onofrio, A. Brandi, S. Felsani, G. Capsoni, G. Drovandi, G. Felici, E. Weitschek, P. Bertolazzi, A. Cattaneo. Gene Expression Biomarkers in the Brain of a Mouse Model for Alzheimer’s Disease: Mining of Microarray Data by Logic Classification and Feature Selection. *Journal of Alzheimer's Disease*, 24(4) 721-738 (2011).
4. E. Weitschek, A. Lo Presti, G. Drovandi, G. Felici, M. Ciccozzi, M. Ciotti, P. Bertolazzi. Human polyomaviruses identification by logic mining techniques. *BMC Virology Journal*, 9:58 (2012).
5. E. Weitschek, G. Fiscon, G. Felici. Supervised DNA Barcodes species classification: analysis, comparisons and results, *BMC BioData Mining*, 7:4 (2014).
6. P. Bertolazzi, G. Felici, P. Festa, G. Fiscon, E. Weitschek. Integer Programming models for Feature Selection: new extensions and a randomized solution algorithm, *European Journal of Operational Research*, 250-389–399, 250 (2016).

Poster 1

**Studying the relationships between bladder cancer and environment
by a genes-miRNAs interaction network approach**

Andrea Polo (1), Anna Crispo (2), Rino Cerino (3), Alfredo Budillon (1), Gennaro Ciliberto (4), Maurizio Montella (2), Susan Costantini (1)

(1)CROM, Istituto Nazionale Tumori “Fondazione G. Pascale” - IRCCS, Napoli, Italia

(2)Unità di Epidemiologia, Istituto Nazionale Tumori “Fondazione G. Pascale” - IRCCS, Napoli, Italia

(3)Istituto Zooprofilattico Sperimentale del Mezzogiorno (IZSM), Portici, Napoli, Italia

(4)Direttore Scientifico, Istituto Nazionale Tumori “Regina Elena”- IRCCS, Roma, Italia

Bladder cancer (BC) is the 9th most common cancer worldwide, the 6th most common cancer in men worldwide and the 4th most common cancer in men in the more developed world.

BC incidence is very high in Campania region of Italy and, in particular, in Naples province. Since the etiology of BC initiation is very multi-factorial, it is always necessary to identify new key genes, miRNAs or metabolites able to improve its early diagnosis and prognosis. Since no detailed information are reported until now on how genes and miRNAs modulated from environmental exposure are correlated between them, our aim was to integrate genes and miRNAs derived from public databases by a computational network approach to identify a subnetwork of genes and miRNAs on which it can be useful to focus further studies.

We extracted from CDT database both the list of the environmental molecules implicated in BC and the genes modulated from them. These molecules were subdivided in four sub-groups like arsenicals, smoking, pollutants and others, and twenty-two genes modulated from all the four sub-groups were selected. These genes resulted to be implicated in the following metabolic pathways: p53 signaling pathway, MAPK signaling pathway, Apoptosis, ErbB signaling pathway, Cell cycle, Glutathione metabolism.

Using Cytoscape program these genes were mapped on Human Molecular Interaction Network, and, to identify the HUB nodes, some statistical analysis were performed such three centrality measures (degree, betweenness and the closeness centrality) and four topological analysis (average characteristic path length, network density, centralization and heterogeneity).

Through these evaluations 15 HUB nodes were identified and a clustering analysis showed the presence in the “bladder cancer and environment network” of four clusters in which five HUB nodes were involved.

Moreover, also the relationships between the circadian genes and HUB genes were evaluated because a recent paper reported that a disturbed function in the cellular clock system represents another mechanism involved in BC progression. Our preliminary

results showed that one HUB node interacts directly with one circadian gene whereas the other HUB nodes correlate with circadian genes in the second order network. Further studies will regard the evaluation of what miRNAs can be able to target 15 HUB nodes to integrate in the network the correlation between genes and miRNAs modulated from the environment and implicated in BC.

Poster 2

Integrator complex: analysis of TCGA omic data across 32 cancer types

Federico A, Rienzo M, Ciccodicola A, Casamassimi A, Costa V.

- 1) Institute of Genetics and Biophysics "Adriano Buzzati Traverso", CNR, Naples.
- 2) Department of Science and Technology, University of Naples "Parthenope".
- 3) Department of Biochemistry, Biophysics, and General Pathology, Second University of Naples.

Integrator complex is a multi-protein transcriptional complex identified only in very recent years. It plays a pivotal role in the regulation of most RNA Polymerase II-dependent genes. This transcriptional complex consists of 14 subunits but its exact structure and protein composition are still largely unknown. Recent studies have highlighted that this complex is crucial for gene expression and regulation in development and in cell differentiation, and it has been demonstrated that alterations (mutations and gene expression changes) of some members of the Integrator family are involved in human diseases, including different types of cancer.

Our work investigates the mutational landscape and the expression of all the members of this gene family across a large panel of human cancers, taking advantage of public datasets of Exome- and RNA-Seq from The Cancer Genome Atlas (TCGA) portal, available for 32 cancer types. Building a computational pipeline based on UNIX shell and R language, we have first identified hotspots of somatic mutations across all cancer types and cancer-specific mutational signatures for these genes. In addition, the analysis of RNA-Seq data shows a wide deregulation of Integrator subunits across several cancer types. Finally, the integration of data deriving from the analysis of somatic mutations and transcriptome data from RNA-Seq in the same patients is still on-going. Preliminary data indicate that genes with somatic mutations in a specific cancer type are highly expressed in tumor tissues.

This pilot study indicates that genes encoding members of the Integrator family are altered in human cancers, suggesting that further targeted studies are needed to ascertain their role in cancer etiology.

Poster 3

Comprehensive analysis of small RNA-Seq data with the iSmaRT toolkit

Memoli D (1), Rinaldi A (1), Panero R (1), Nassa G (1,2), Rizzo F (1), Ravo M (1), Tarallo R (1), Weisz A (1), Giurato G (1,2)

(1) Laboratory of Molecular Medicine and Genomics, Department of Medicine, Surgery and Dentistry 'Scuola Medica Salernitana', University of Salerno, Via S. Allende 1, 84081 Baronissi (SA), Italy

(2) Genomix4Life, University of Salerno, Via S. Allende 1, 84081 Baronissi (SA), Italy

The interest in investigating the involvement of small non-coding RNAs (sncRNAs) is increasing, due to the pleiotropic roles these molecules exert in many biological contexts. While several methods and tools, such as for example iMir [1], are available to study microRNAs (miRNAs), only few focus on novel classes of sncRNAs, in particular on PIWI-interacting RNAs (piRNAs). In order to satisfy the growing demand for intuitive and efficient data exploration tools useful to this end, we developed iSmaRT (integrative Small RNA Tool-kit) [2], integrating specific bioinformatics softwares and statistical approaches for a comprehensive analysis of smallRNA-Seq data. The interconnection of several analytical features for identification and comprehensive downstream analyses on sncRNAs makes iSmaRT very flexible and easy to use, also to users with limited bioinformatics background. iSmaRT is written in Python and developed in a virtual machine, integrating several third-party softwares widely used for the analysis of smallRNA-Seq with own algorithms. The Graphical User Interface (GUI) interconnects all these tools estranging the users from using R and bash command manually. Specifically for piRNAs we have developed two modules that allow the users to find piRNA targets and to perform the prediction of novel one. Tests on different smallRNA-Seq datasets indicate that iSmaRT allows a comprehensive analysis of different classes of sncRNAs in a rapid and user friendly way, demonstrating its usefulness for extracting novel information related to sncRNAs in different experimental settings.

iSmaRT is freely available on the web at <ftp://labmedmolge-1.unisa.it> (User: iSmart - Password: password).

[1] G. Giurato et al., BMC Bioinformatics 14:362, 2013

[2] R. Panero et al., Bioinformatics 2016. In press

Poster 4

**Investigating the interaction between transglutaminase 2 and fibronectin:
an in vitro and in silico study**

Ida Autiero (1), Maria Felicia Soluri (2), Luigi Cavallo (1), Daniele Sblattero (2) and Romina Oliva (3)

(1) Kaust Catalysis Center, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia.

(2) Department of Life Sciences, University of Trieste.

(3) Department of Sciences and Technologies, University “Parthenope” of Napoli.

In the extracellular environment, the enzyme transglutaminase 2 (TG2) is involved in cell–matrix interactions through association with the extracellular matrix protein, fibronectin (FN). The 45 kDa gelatin-binding domain of FN (45FN) is responsible for the binding to TG2. Previous studies have shown that: i) the FN-binding site of TG2 is located in the N-terminal domains of the enzyme, although with conflicting results regarding the specific residues involved in the interaction, and ii) TG2 can be bound to FN both in its open and closed conformations.

Experimental results we present herein confirm the two points above. In addition, we show that the FN domain 8 is sufficient to achieve an efficient binding to TG2, although a minimum fragment containing FN domains 7 to 9 is required for preserving the biological activity in cells. To model the interaction between FN and TG2 at an atomic level, we initially applied an unbiased approach, using six different docking servers to predict the 3D structure of the protein-protein complex, without driving the docking to any specific residue. Then, we collected all the obtained models and, on the basis of the “consensus” experimental information, we filtered out all the models where FN was bound to the TG2 C-terminal domains and/or not compatible with either the closed/open TG2 state. Upon clustering the remaining models [1], we identified two highly recurrent solutions, both pointing to a binding interface on TG2 recently suggested by Sollid et al., on the basis of hydrogen/deuterium exchange and mass spectrometry experiments [2]. Details on the obtained models will be presented.

1. Chermak, E., De Donato, R., Lensink M.F., Petta, A., Serra, L., Scarano, V., Cavallo L., Oliva R. (2016) Introducing a Clustering Step in a Consensus Approach for the Scoring of Protein-Protein Docking Models, PLoS One 11:e0166460.

2. Cardoso, I., Østerlund, E.C., Stamnaes, J., Iversen, R., Andersen, J.T., Jørgensen, T.J., Sollid, L.M. (2016) Dissecting the interaction between transglutaminase 2 and fibronectin, Amino Acids, DOI: 10.1007/s00726-016-2296-y.

Poster 5

Transduction of allosteric signaling in the human CXCR3/CXCL11 complex by a combined dynamics-networking approach.

Giovanni Colonna (1), Andrea Polo (2), Susan Costantini (2)

(1) Centro di Informatica Medica, Azienda Ospedaliera Universitaria della Università degli Studi della Campania "Luigi Vanvitelli"

(2) CROM, Istituto Nazionale tumori "Fondazione G. Pascale", IRCCS, Napoli

The seven-helices trans-membrane receptor of human chemokines, CXCR3, that are studying from some years [1-4], belongs to the class of receptors that bind a cytoplasmic "G-protein" following the extra-cellular binding of one of its natural ligands, the human chemokine CXCL11. One of the issues currently under intense studies is the understanding of the allosteric mechanism of transduction through receptors of a molecular signal that, reaching the C-terminal region, activates the binding with the G-protein, thus triggering the biological action at cytoplasmic level. This phenomenon is totally dynamic with the receptor in the membrane; therefore, X-ray and NMR are not of great help to solve a phenomenon that proceeds by sequential conformational changes through lowest energy conformers.

In this work we have modeled in the membrane and in the presence of water the receptor alone and the CXCR3 / CXCL11 complex by docking. The two systems were subjected to molecular dynamics parameterizing the dynamics not so much with the usual parameters (RMSD, radius of gyration, etc.) but, rather, with those used to assess the residue-residue interactions in the protein networks. Basically, we have considered that the two models are at equilibrium during the dynamics only when the total number of residue-residue interactions does not change anymore. The rationale of this approach is based on the fact that the signaling propagation determines a specific pathway of allosteric conformational changes that, proceeding along the molecule, from side to side, as a consequence they determine changes in the residue-residue interactions. An analysis of "residue-residue interaction network" (RIN) of models before and after dynamics sees exactly how individual residues interact with each other by changing number and type of interactions up to the equilibrium state.

Comparing with this same approach the two models, CXCR3 alone and complexed, the allosteric wave passes through the receptor proceeding with a well-defined conformational pathway that starting from the N-terminal involves the helices TM1 and TM7, quite indirectly the loops ICL1 and ICL2 and some residues of the C-terminal. The analysis of the two dynamics (clustering, vector projections, transition rates, etc.) has allowed us to make an assessment of the energetics of the signal progression that dynamically involves the lowest energy structures we previously suggested.

- [1] Trotta T, Costantini S, Colonna G. Modelling of the membrane receptor CXCR3 and its complexes with CXCL9, CXCL10 and CXCL11 chemokines: putative target for new drug design. *Mol Immunol*. 2009 Dec;47(2-3):332-9.
- [2] Palladino P, Portella L, Colonna G, Raucci R, Saviano G, Rossi F, Napolitano M, Scala S, Castello G, Costantini S. The N-terminal region of CXCL11 as structural template for CXCR3 molecular recognition: synthesis, conformational analysis, and binding studies. *Chem Biol Drug Des*. 2012 Aug;80(2):254-65.
- [3] Costantini S, Raucci R, De Vero T, Castello G, Colonna G. Common structural interactions between the receptors CXCR3, CXCR4 and CXCR7 complexed with their natural ligands, CXCL11 and CXCL12, by a modeling approach. *Cytokine*. 2013 Oct;64(1):316-21.
- [4] Raucci R, Colonna G, Giovane A, Castello G, Costantini S. N-terminal region of human chemokine receptor CXCR3: Structural analysis of CXCR3(1-48) by experimental and computational studies. *Biochim Biophys Acta*. 2014 Oct;1844(10):1868-80.

Poster 6

A comparison of predictive algorithms on a novel SNP-panel for ancestry determination in forensic genetics.

Daniela Evangelista (1), Marina Piccirillo (1), Gianluca Peri (2), Greg Hampikian (2), Mario R. Guarracino (1)

(1) Laboratory for Genomics, Transcriptomics and Proteomics (LAB-GTP), High Performance Computing and Networking Institute (ICAR), National Research Council of Italy (CNR), Via Pietro Castellino, 111, Napoli, Italy

(2) Department of Biological Sciences and Department of Criminal Justice Boise State University Dr.,1910, Boise, USA

To support criminal investigations, many panels of Single Nucleotide Polymorphisms (SNPs) have been developed to provide valuable BiogeoGraphical Ancestry (BGA) information for evidentiary DNA samples not matching any record in available investigative databases. Starting from 2 well-established ancestry informative SNPpanels, using data of genetic variability extracted from the 1000 Genomes Project database, we created a unique panel consisting of a total of 622 SNPs and 10 attributes. The aim of this study is to evaluate the predictive and discriminative power of our new SNP-panel belonging to four super-populations in a world wide contest: African, European, Asiatic and Native American, as well as, to perform a direct comparison of ancestry prediction capability among 11 different machine learning algorithms. The obtained results are gathered, organized and shown in forenSNP, a comprehensive web resource available at: www-labgtp.na.icar.cnr.it/forenSNP

Poster 7

Insight into the structural features of vertebrate myoglobin family

Sara Ragucci, Antonella MA Di Giuseppe, Nicola Landi and Antimo Di Maro

Dip. di Scienze e Tecnologie Ambientali, Biologiche e Farmaceutiche, Università degli Studi della Campania “Luigi Vanvitelli”

Myoglobin (Mb) is a monomeric hemoprotein that binds oxygen in oxidative skeletal muscle and cardiac myocytes. Mb facilitates oxygen diffusion to the mitochondria of aerobic muscle, and its storage for metabolic respiration. Structurally, despite differences in primary amino acid sequences, Mbs from all species are predicted to fold into the same tertiary structure referred to as the ‘globin fold’. Despite the divergence of myoglobin sequences, the ‘globin fold’ is preserved and few amino acid residues such as the proximal and the distal histidinyll residues are highly conserved. Nevertheless, the intrinsic stability of this fold is coupled to conformational changes enabling to alter oxygen affinity in Mbs from different organisms. Moreover, the physico-chemical peculiar properties of Mbs and its easy availability for different sources made this protein a formidable archetype to investigate and clarify the structure/function relationships in protein science.

In this work, we have extensively used database search techniques to shed light on this important globin family, using the amino acid sequences from 298 vertebrate, classified in five clades in correspondence with five classes of gnathostomes: Chondrichthyes (cartilaginous fishes), Actinopterygii (ray-finned fishes), Reptilia, Mammalia and Aves. Forty-one high conserved residues were identified and seven of them were invariants [E18, G25, F43, V68, L72, H93 (proximal histidine) and H97]. E18 is the only invariant amino acid residue located out of the heme-pocket and X-cavities playing a role in interaction between the A and E-helices. A comparative analysis of several parameters related to amino acid composition shows an increase of average mass, accessible surface area and volume per residue from Actinopterygii to Mammalia and Aves. This may be due to an increased number of bulky residues reducing the non-specific cavities volume and thus improving the oxygen flow between the heme site and the outside of the protein.

Thus, this study confirms that despite sharing a similar fold, our analysis highlights the plasticity of vertebrate myoglobin, which allows adaptability of this protein under different oxygen demand to which vertebrates are subjected.

Poster 8

Drug discovery by integration of pharmacophore modeling, virtual screening and biological evaluation by means of bioinformatics resources

S. Dotolo (1,2), A. Facchiano (1)

(1) Institute of Food Science (ISA-CNR), Via Roma 64, Avellino (Italy)

(2) Department of Biochemistry, Biophysics and General Pathology, Università degli Studi della Campania “Luigi Vanvitelli” Via de Crecchio 7, 80138 Naples (Italy)

serena.dotolo@isa.cnr.it, angelo.facchiano@isa.cnr.it

Drug discovery is a step-by-step process very important in biopharmaceutical field. We are interested in identifying new investigational *drug-likes* as potential inhibitors of determinate biological-therapeutic targets, trying to decrease the side effects and to safeguard the human health [1-2]. However, it is a long and very expensive process [3]. Therefore, we are using a new computational strategy, based on *Pharmacophore modeling*, to select bioactive substances (natural or synthetic), through the integration of bioinformatics online tools and local resource and platforms, in order to include into the strategy also knowledge from high-throughput studies, for new potential *lead compounds* generation-optimization, trying to accelerate the early phase of the drug development process [4].

The protocol of this new computational strategy is characterized by a multi-step design focused on: 1) *screening in RCSB-PDB for a crystal structure of a specific biological target*, suitable for the following steps; 2) *pharmacophore modeling and virtual computational screening*, by using public domain databases of bioactive compounds, as the ZINC¹² database [5], in order to find a promising molecule that could become a new potential medicine. 3) *molecular and biological evaluation*, to check the compounds selected by virtual screening, for their biological properties through public databases, as PubChem Compound, SciFinder, and Chemicalize to trace their origin and underline their most important physical-chemical features, PathPred (an enzyme-catalyzed metabolic pathway predictor server) to highlight and identify their biosynthetic-metabolic pathways and investigating the biotransformation of best candidates, analyzing their metabolites and their potential biological activity. Moreover, ADMET/toxicity predictor server applying the Lipinski-Veber filter are used to calculate the bioavailability the ADMET/toxicity properties.

After this check, only molecules with good bioavailability, good predicted activity and good ADMET properties are considered as hits compounds or *drug-likes* to direct the design of next experimental assays [6].

Finally, the lead compounds selected are analyzed through *molecular dynamics simulations*. 4) *simulations of molecular dynamics on the best lead compounds*, to investigate atomic details of protein-compound molecular interactions in different conditions (different organic solutions, organisms and systems).

REFERENCES

- [1] Dubey A, Facchiano A, Ramteke PW, Marabotti A. "In silico approach to find chymase inhibitors among biogenic compounds." *Future Med Chem.* 2016; 8(8):841-51
- [2] Dubey A, Marabotti A, Ramteke PW, Facchiano A. "Interaction of human chymase with ginkgolides, terpene trilactones of Ginkgo biloba investigated by molecular docking simulations." *Biochem Biophys Res Commun.* 2016; 473(2):449-54.
- [3] Katara P. "Role of bioinformatics and pharmacogenomics in drug discovery and development process". *Netw Model Anal Health Inform Bioinforma* 2013; 2: 225-230.
- [4] Sunseri J. and Koes D. R. "Pharmit: Interactive Exploration of Chemical Space". *Nucl. Acids Res.* 2016; 44(W1): W442-448.
- [5] Irwin J.J. and Shoichet B.K. "ZINC- A free database of Commercially Available Compounds for Virtual Screening". *J.Chem.Inf.Model.* 2005; 45: 177-182.
- [6] Kaserer T., Beck K. R., Akram M., Odermatt A., Schuster D. "Pharmacophore Models and Pharmacophore-Based Virtual Screening: Concepts and Application Exemplified on Hydroxysteroid Dehydrogenases". *Molecules* 2015; 20: 22799–22832.