POSTER LIST
ORDERED ALPHABETICALLY BY POSTER TITLE
GROUPED BY THEME/TRACK

THEME/TRACK: PROTEINS

Poster numbers: P_Pr001 - 080 Application posters: *P_Pr001 - 009*

Poster number	EasyChair number	Author list	Presenting author	Title	Abstract	Theme/track	Topics
APPLICATION POSTERS WITHIN PROTEINS THEME							
P_Pr001	*674*	*Fatemeh Abbasi, Changiz Eslahchi and Reza Hassanzadeh*	*Fatemeh Abbasi*	*A GRAPH THEORETICAL APROACH FOR DRUG TARGET PREDICTION*	*Motivation: The discovery of novel drug targets is a significant challenge in drug development. Many of the currently known drug targets are functionally pleiotropic and involved in multiple pathologies. Several of them are exploited for treating multiple diseases, which highlights the need for methods to reliably reposition drug targets to new indications. So, the identification of interactions between drugs and target proteins is a key area in genomic drug discovery. Therefore, there is a strong incentive to develop new methods capable of detecting these potential drug–target interactions efficiently. Computational methods for novel drug target predictions can greatly reduce time and costs compared with experimental methods.Results: In this work, we present a network based computational approach for novel drug and target association predictions. More specifically, a heterogeneous drug-target graph, which incorporates known drug-target interactions, is first constructed. Based on this graph, drug-drug and target-target similarities, a novel graph based inference method is introduced. Compared with two state of the art methods, 10 fold cross-validation and jackknife results on different data sets, involving targets of enzyme, ion channel, GPCR, nuclear receptor and complete DrugBank indicate that the proposed method can greatly improve novel target predictions.*	*Proteins/ Application poster*	*Application Fundamental*
P_Pr002	*673*	*Changiz Eslahchi, Ali Madi and Changiz Eslahchi*	*Changiz Eslahchi*	*Discovering overlapped protein complexes from weighted PPI networks by removing inter-module hubs*	*Motivation: Detecting known and predicting undiscovered protein complexes from protein-protein interaction (PPI) networks helps us to understand principles of cellular organization and their functions. Nevertheless, extraction of protein complexes from PPI network isn't an easy task. Two major constrains are high noise level and ignoring occurrence time of different interactions in PPI network.Results: An efficient algorithm (IMHRC) is developed based on inter-module hub removal in the weighted PPI network which can detect overlapped complexes. IMHRC by removing some of the inter-module hubs and module hubs, eliminates a meaningful percentage of noise in our dataset and indirectly consider difference occurrence time of the PPI in our network. After removing hubs, some proteins are considered as seeds. Each seed creates a primary cluster. Then removed module hubs are added to the resulting clusters based on the amount of their interactions with other proteins in the clusters. Clusters are then merged based on their overlaps. Consequently, the performance of the IMHRC is evaluated on several benchmark datasets and the results are compared with other state-of-the-art models. The protein complexes that discovered by IMHRC method significantly match with the real data and much better than other methods.*	*Proteins/ Application poster*	*Application Fundamental*
P_Pr004	*847*	*Thomas Kemmer and Andreas Hildebrandt*	*Thomas Kemmer*	*Efficient nonlocal electrostatics computations for proteins using the Julia programming language*	*Electrostatic interactions are a major contributor to protein-protein and protein-ligand interactions. In contrast to other molecular interaction components, they can be significant over medium to long distances and are thus crucial for molecular visibility. Research areas such as rational drug design require accurate estimates of potentials and free energies influenced by electrostatics. One major challenge in this context, however, is the treatment of the solvent the molecules are immersed in, i.e., water in a biological context. Strong simplifications of the structure of such polarizable and highly structured solvents are commonplace to achieve the required computational efficiency, but invariably lead to inaccuracies.Here, we present efficient protein electrostatics computations in a single and easily extensible software package for the cross-platform and open-source Julia programming language. By modelling water in an implicit but nonlocal fashion, we account for correlation of molecular polarization due to the water network around the solute and sustain accuracy without suffering from infeasible runtimes as compared to the explicit case. Our package contains implementations for our own Boundary Element (BEM) solver as well as a reference Finite Element (FEM) solver, both profiting from the good base performance of the Julia language, which can achieve runtimes comparable to C. Additionally, Julia's native and non-native interoperability with other languages such as C, Fortran, R, and Python allows for easy incorporation of our package into existing pipelines.*	*Proteins/ Application poster*	*Application*
P_Pr005	*472*	*Saba Ferdous and Andrew Martin*	*Saba Ferdous*	*Exploration of conformational B-cell epitopes: components to peptide-based vaccines*	*Peptide vaccines have many potential advantages including low cost, lack of need for cold-chain storage and safety. However, it is well known that approximately 90% of B-cell Epitopes (BCEs) are discontinuous in nature making it difficult to mimic them for creating vaccines. We have analyzed the discontinuity of B-cell epitopes by defining extended 'regions' (R, consisting of at least 3 antibody-contacting residues each separated by <= 3 residues) and small fragments (F, antibody-contacting residues that do not satisfy the requirements for a region). Secondly, we have classified region shape as linear, curved or folded. Furthermore, by using molecular dynamics, we have studied mutations in linear and folded (two alpha helices or beta strands connected by hairpin loop) regions that stabilize their conformation: end capping, mutations of hydrophobics (non contacting residues of an epitope) to alanine and glutamine, disulphide stapling and cyclization. We have explored mutations in five linear and five folded epitopes with up to 20 mutant for each of the epitopes. Moreover, to confirm the stability of a stable mutant in the presence of an antibody, it has been simulation with antibody. The stablised epitope mimetics (mutant) will be tested experimentally to check their possibility to use as immunogens for peptide vaccine design.*	*Proteins/ Application poster*	*Application Health*
P_Pr006	*326*	*Anoosha Paruchuri, Huang L-T, Sakthivel R, Karunagaran D and Michael Gromiha M*	*Anoosha Paruchuri*	*Exploring preferred amino acid mutations in cancer and discriminating driver and passenger mutations in Epidermal Growth Factor Receptor*	*Cancer is one of the leading causes of death worldwide. Huge number of somatic mutations get accumulated during cancer development, among which contributes to tumor progression are known as "driver" mutations, whereas most of them are functionally neutral known as "passenger" mutations. Hence, discriminating these mutations has been an active field in cancer research. In this study, we have systematically analysed the effect of these mutations at protein level in 41 different cancer types from COSMIC database on different perspectives: (i) Preference of residues at the mutant positions (ii) Probability of substitutions (iii) Influence of neighbouring residues (iv) Distribution of driver and passenger mutations around hotspot sites and (v) Distribution of silent and missense substitutions. This study reveals the variation of mutations at protein level in different cancer types and their preferences in cancer genes and provides new insights for understanding cancer mutations and drug development. Further, considering the importance of EGFR (Epidermal Growth Factor Receptor) protein based on the number of observed missense mutations in cancer, we have developed a reliable classification model for discriminating driver and passenger mutations in this protein. We grouped the mutations based on secondary structure and accessible surface area and acheived an overall classification accuracy of 80.2%, 81.9%, 77.9% and 75.14% for helix, strand, coil buried and exposed mutants, respectively. We have screened all possible missense mutations in EGFR and suggested probable driver and passenger mutations, which would help in the development of mutation specific drugs for cancer treatment.*	*Proteins/ Application poster*	*Application*
P_Pr007	*368*	*Rakesh Kumar Meena, Sayane Shome and Sanket Thakur*	*Rakesh Kumar Meena*	*In silico prediction of Okra (Abelmoschus esculentus L) encoded micro-RNAs targets, Structure prediction and Molecular docking studies for Okra yellow Vein Mosaic virus.*	*Begomovirus associated symptoms were observed in several Abelmoschus esculentus plants growing in crop fields in India as well as whole world. Protein sequence of the viral coat protein from the yellow vein mosaic virus was collected from NCBI protein database [Accession ID: NP_579972].The nucleotide sequence and the coordinates of an Okra leaf isolate was obtained from NCBI Nucleotide database [Accession ID: KC404826]. The nucleotide sequence was then subjected to sequence search in MiRbase which utilizes BLASTN algorithm to find candidate miRNAs deposited from the database. The miRNA determined in the nucleotide sequence[Accession ID : MI0027065] lies in the interspacer region which further supports our claim for the miRNA candidate for the analysis.3-dimensional coordinates for the viral coat protein and the miRNA candidate was predicted by Modeller software and I-Tasser server (http://zhanglab.ccmb.med.umich.edu/I-TASSER/) . The best suitable structure was determined by validation using SAVS server and observed more than 93% residues in core region. Structural annotation and cavity prediction was carried out by PSI-Phred server and Castp server respectively. The 3-dimensional structure of miRNA candidate was predicted using Chimera software. Molecular docking of the viral coat protein with miRNA candidate was carried out via Autodock4.2 software. Grid coordinates was determined by Autogrid 4 and Lamarckian Genetic algorithm was used for the docking process. Interacting residues participating in the molecular docking was visualized by Chimera and Ligplot+ software. The computational study comprising molecular binding of OYVMV coat protein with miRNA from Okra leaf isolate shows promising results which can be replicated in experimental studies to devise novel therapeutic strategy to treat Okra yellow vein mosaic viral disease.*	*Proteins/ Application poster*	*Application Biotechnology*
P_Pr008	*858*	*Pooya Zakeri, Jaak Simm, Adam Arany, Forough Amini, Mehdi Sadeghi and Yves Moreau*	*Pooya Zakeri*	*Protein Fold Recognition Using Matrix Factorization Technique*	*Most of protein fold predictor machines only cover less than 30 folds, which is far less than protein folds have been identified. Moreover, the typical approaches proposed for protein fold recognition often neglect the relationship between protein folds. These motivate us to formulate the protein fold recognition as a factorization of an incompletely filled binary protein-fold-matrix where the objective is to predict unknown values. Protein fold recognition database such as SCOP can be seen as an incomplete matrix (M*N) where each row is a protein and each column is a protein fold. Then, the SCOP matrix can be modeled by the two smaller matrices P and F, which, when multiplied, approximately reconstruct the SCOP matrix. We propose an extended version of the Bayesian probabilistic matrix factorization [1] with the added advantage of working with multiple protein features as side sources for completing protein-fold-matrix. Accordingly, two sequence-based protein features, including the predicted secondary structure and information extracted directly from position-specific scoring matrices, are incorporated into the proposed factorization model as side information.In order to validate our models in a more realistic task setting, we develop a prospective benchmark, based on the latest version of the SCOP database, which covers about 200 protein folds. The experimental results on our unbiased benchmark show that our proposed model can effectively improve the accuracy of the state-of-the-art protein fold predictors such as GeoFold [2].[1] doi: 10.1093/bioinformatics/btu118.[2] doi: 10.1145/1390156.1390267.*	*Proteins/ Application poster*	*Application Fundamental*
P_Pr009	*606*	*Dhoha Triki, Telli Billot, Benoit Visseaux, Diane Descamps, Anne-Claude Camproux andLeslie Regad*	*Dhoha Triki*	*Study of natural resistance mechanisms of HIV protease-2 (PR2) against protease inhibitors (PI)*	*The therapeutic arsenal against the HIV of type 2 (HIV-2) corresponds to antiretroviral drugs developed for HIV-1. HIV-2 is naturally resistant to some of these drugs. It is therefore important to find new drugs against HIV-2. A solution is to develop specific molecules inhibiting HIV-2 protease (PR2), an enzyme involved in the maturation of virus proteins (Brower et al. 2008). Understand what factors contribute to the efficiency of inhibitors for HIV-1 protease (PR1) and absent from the PR2 can help to improve the PR2 inhibitor design.In this study, we compared a set of 36 structures of PR1 and PR2. They exhibit 48% of sequence identity. Mutations between PR1 and PR2 are primarily located on the elbows regions and few mutations are located in the PI-binding site. We analyzed the effects of these mutations on PR2 structure. First we observed that these mutations seem to modify the PR2 flexibility: PR2 structures have on average higher Bfactor values, meaning PR2 structures are more flexible than PR1. We then noted that these mutations have an effect on the properties of PI-binding sites: those extracted from PR2 structures are less hydrophobic, smaller and more polar than those of PR1. Finally, we observed that these mutations modify PR interface properties: PR2 dimer structures exhibit a lesser energetic stability than PR1 interfaces. To conclude, our study showed that mutations between PR1 and PR2 have important effects on PR. Molecular dynamics simulations could be used to understand the effect of these mutations on the PI-binding mode.*	*Proteins/ Application poster*	*Application Health*
OTHER POSTERS WITHIN PROTEINS THEME							
P_Pr010	*389*	*Patrick Löffler, Samuel Schmitz, Enrico Hupfeld and Rainer Merkl*	*Patrick Löffler*	*A Modular Framework to Extend Rosetta Protocols with Multistate Design*	*Computational protein design (CPD) is a powerful technique to design novel proteins. Many CPD objectives such as design on backbone ensembles, multi-specificity design and the integration of negative design demand the simultaneous optimization of multiple design states. Rosetta is a popular software suite to study and design proteins. Rosetta's protocols consist of specific procedures and a fine-tuned set of parameters to carry out a given task. An example is the use of specific sequence design cycles and catalytic constraints in the enzyme design protocol. At present, the multistate design implementation of Rosetta is a generic approach lacking options to fine tune the calculations in the same manner as specialized single state protocols.We have developed a framework for CPD that integrates multistate design in existing Rosetta protocols while preserving the protocol's original functionality. Our framework consists of two, easily exchangeable components: i) The optimizer searches the sequence space and ii) the evaluator scores the sequences according to the given design task. Currently, we utilize Rosetta's genetic algorithm as an optimizer; protocols for enzyme design and protein-protein interface design serve as evaluators. However, due to the modularity of both components, multistate functionality can be transferred to arbitrary Rosetta applications with little effort. We have benchmarked the above two applications on two datasets consisting of conformational ensembles and achieve an 18 percent performance improvement over conventional methods. As a proof of concept, we have applied our framework to computationally design retro-aldolases which are currently subject to biochemical characterization.*	Proteins poster	Biotechnology
P_Pr011	*854*	*Dina Cramer, Luis Serrano and Martin H Schaefer*	*Martin H Schaefer*	*A network of epigenetic modifiers and DNA repair genes controls tissue-specific copy number alteration preference*	*Copy number alterations (CNAs) show a large variability in their number, length and position over cancer types. This variability is clinically relevant as both the amount and length of CNAs (as well as the identity of the affected genes) have a strong impact on patient survival. However, the sources of this variability are not known. We aim to identify genetic and epigenetic factors that contribute to this variability.Analyzing patient data from The Cancer Genome Atlas (TCGA), we have identified proteins that tend to be mutated in samples having few or many CNAs, which we term CONIM proteins (COpy Number Instability Modulators). CONIM proteins cluster into a densely connected subnetwork of physical interactions and many of them are epigenetic modifiers. Therefore, we investigate how the epigenome of the tissue-of-origin influences the position of CNA breakpoint regions and the properties of the resulting CNAs. We find that the presence of heterochromatin in the tissue-of-origin contributes to the recurrence of CNAs in the respective cancer type. We show that these epigenetic states also impact the length of the resulting CNAs, elucidating differences in the mechanisms underlying CNA generation.Therefore, we demonstrate how both the tissue-of-origin epigenome organization and a newly identified class of cancer genes affect the variability of CNA number over patients and cancer types.*	Proteins poster	Health
P_Pr012	*483*	*Isaure Chauvot de Beauchene, Sjoerd De Vries and Martin Zacharias*	*Isaure Chauvot de Beauchene*	*A new fragment-based docking approach to model protein-bound ssRNA from sequence.*	*Protein-RNA recognition supports many cellular functions. Abnormal protein-RNA interactions are crucial therapeutic targets in e.g. neurodegenerative diseases and RNA viruses infections. Moreover, synthetic RNA aptamers can be used as protein modulators. The rational design of either aptamers or RNA-protein interaction inhibitors requires atomistic description of protein-RNA complexes. Yet their experimental resolution is arduous, and protein-RNA computational docking is hampered by the high flexibility of RNA single-stranded regions, which mostly provides recognition specificity. The lack of methodology for modeling ssRNA limits all protein-RNA docking methods [2].We developed an original fragment-based approach, predicting ssRNA-protein complexes structure from protein structure and RNA sequence: We (i) cut the RNA sequence in overlapping trinucleotides, represented by sequence-specific ensembles of conformers that we built from known protein-RNA structures; (ii) dock each ensemble on the protein; (iii) select the spatially compatible poses; (iv) assemble them in a realistic conformation.Moreover, we developed and validated an RNA-protein contact predictor, based on statistical analysis of known complexes, which provides (optional) starting points for fragments docking.We applied them on ten complexes with various ssRNA sequences (6-11 nucleotides) and RNA-recognition domains. Without predicting specific contacts, we can identify the RNA binding site more accurately than existing methods [3]. Based on predicting 3-4 contacts, the method allows modeling of bound ssRNA within 1-2Å RMSD [1]. Such crystallographic-like precision, not reached so far , reveals a methodological breakthrough in RNA-protein docking.[1] Chauvot-de-Beauchene et al (2016) NAR 44(10):4565-4580[2] Fulle, Gohlke (2010) JMR 23(2):220-231[3] Chauvot-de-Beauchene et al (2016) PloS.Comput.Biol. 12(1):e1004697*	Proteins poster	Biotechnology

P_Pr013	824	Mark Wass, Sarah Jeanfavre, Michael Coghlan, Martin Ridout, Anthony Baines and Michael Geeves	Mark Wass	Adaptation of mammalian myosin II sequences to body mass	The speed of muscle contraction is related to body size; muscles in larger species contract at a slower rate. We investigated the evolution of twelve myosin II isoforms to identify any adapted to increasing body mass. β-myosin head domain had the greatest rate of sequence divergence (0.05% per Myr) and was the only domain where sequence divergence correlated with body mass (0.091% divergence per log mass unit). β-myosin is abundant in cardiac ventricle and slow skeletal muscle. We propose that β-myosin has adapted to enable slower heart beating and contraction of slow skeletal muscle as body mass increased. Additionally, for eight of the twelve myosins, the ratio of divergence in the head:tail domains was significantly different, ranging from 3:1(β-myosin) to <1:2(extraocluar, non-muscle A and embryonic myosin). Our data provide new insights into the evolution of myosin function and indicate distinct evolutionary pressures on head and tail domains in individual isoforms.	Proteins poster	Fundamental
P_Pr014	452	Michal Burdukiewicz, Piotr Sobczyk, Stefan Rödiger, Paweł Mackiewicz and Małgorzata Kotulska	Michal Burdukiewicz	AmyloGram:a novel predictor of amyloidogenicity	Amyloids are proteins associated with the number of clinical disorders (e.g., Alzheimer's, Creutzfeldt-Jakob's and Huntington's diseases). Despite their diversity, all amyloid proteins can undergo aggregation initiated by 6- to 15-residue segments called hot spots. Henceforth, amyloids form unique and often zipper-like β-structures, which can turn out harmful.To find patterns defining the hot-spots, we trained predictors of amyloidogenicity based on random forests using n-grams extracted from amyloidogenic and non-amyloidogenic peptides collected in the AmyLoad database.Since the amyloidogenicity may not depend on the exact sequence of amino acids but on more general properties of amino acids in the sequence, we constructed 524 284 reduced amino acid alphabets of different lengths (three to six letters) based on all possible combinations of the handpicked physicochemical properties of the amino acids. The cross-validation of predictors employing the different alphabets revealed the best-performing alphabet with the length of 6 amino acid residues. During analysis we found also 65 n-grams that are the most relevant to the discrimination of amyloid and non-amyloid sequences of which 15 were confirmed experimentally elsewhere.The best-performing predictor, AmyloGram, was benchmarked against the most popular tools for amyloid peptides detection using an external dataset. It has obtained the highest values of performance measures (AUC: 0.90, MCC: 0.63).AmyloGram is available as a web-server: www.smorfland.uni.wroc.pl/amylogram/.	Proteins poster	Health
P_Pr016	531	Dhoha Triki, Mario Cano Contreras, Delphine Flatters, Benoit Visseaux, Diane Descamps, Anne-Claude Camproux andLeslie Regad	Leslie Regad	Analysis of the HIV-2 protease deformation involved by inhibitor binding	HIV-2 is a retrovirus discovered a few years after HIV-1. HIV-2 infections are restricted mainly to West Africa and to some European countries (Valadas et al., 2009 ;Brunet S, et al. 2008). The HIV-1 and HIV-2 genomes differ by about 50% at the nucleotide level. Such differences may be correlated with differential responses to some antiretrovirals such as some protease inhibitors (PIs) (Poveda E et al., 2005, Ren J, et al., 2002). It is necessary to develop new therapeutic molecules specific to HIV-2. One approach is based on the identification of new molecules inhibiting the HIV-2 protease (PR2), a protein involved in HIV-2 protein maturation. To do so, it is important to understand which features are contributing to the PI selectivity and efficiency for the HIV-1 protease (PR1) and absent in PR2. The understanding of the interaction mode between antiretroviral drugs with the PR2 and the PR2 structural deformation implied by the inhibitor binding can help to this task.In this study, we first compared the inhibitor-binding pockets extracted from 19 Xray structures of PR2 (apo and holo forms). In a second step, we analysed the PR2 plasticity using SA-conf tool. This tool analyzes the structural plasticity of a target by comparing the local structures of its different conformations. SA-conf allowed us to highlight the PR2 structural variable regions putatively involved by the PI-binding. This study allowed us to detect residues important for the inhibitor binding in PR2 and to better understand the PR2 deformation implies by the inhibitor-binding.	Proteins poster	Health
P_Pr017	651	Galo Ezequiel Balatti, M. Florencia Martini and Monica Pickholz	Galo Ezequiel Balatti	Antimicrobial peptides mechanisms of membrane lysis and permeation by computer simulations	Antimicrobial peptides (AMPs) are part of the innate immune system, attaching and inserting to the lipidic membranes of external agents among bacteria, fungi, viruses and eukaryotic parasites and killing the cells through a membrane permeation effect.Nevertheles, their molecular mechanisms are not well-known and three different leakage pathways was proposed: the "barrel-stave", the "carpet" or the "toroidal-pore" models. Among AMPs, two peptides obtained from Australian tree frogs, the Aurein 1.2 and the Maculatin 1.2 are proposed as AMPs with different leakage pathways. Here, we carried out extensive Molecular Dynamics (MD) simulations to study the peptide interactions with lipid structures in order to shed light into these mechanisms.. We have used a coarse grain (CG) model within the MARTINI force field[1]. Three simulation replicates were performed, looking to the self-assembly of 1000 lipids (2-oleoyl-1-palmitoyl-sn-glycero-3-phosphocholine, POPC) in the presence of the peptides were performed. Furthermore, we simulate both peptides in a presence of a pre-equilibrated bilayer from two different initial configurations: aqueous phase and inside the bilayer.The simulations results showed two different pathways on the membrane leakage, in good agreement with experimental observations.[2] While Maculatin can form a pore maintaining the structure of the bilayer, Aurein causes the total membrane destabilization and disintegration. A better understanding of AMPs molecular behavior can aim the development of new antimicrobials drugs.[1] X. Periole, S.J. Marrink. Methods in molecular biology 925 (2013) 533-565[2] E.E. Ambroggio et al Biophysical Journal 89 (2005) 1874–1881	Proteins poster	Health
P_Pr018	693	Maria Katsantoni, Tjaart de Beer and Torsten Schwede	Maria Katsantoni	Assessing functional conservation in alternative splice forms	In 75% of human genes, alternative splicing gives rise to more than one transcript per gene. However, little is known about the functional significance these alternative products have. Thanks to RNA-seq technology, human transcriptome data are constantly increasing, which gives a better view of how alternative transcripts are expressed under different conditions e.g. normal and cancer tissue data.In this work we focus on the alternative protein-coding transcripts and what their functional importance may be on the protein level. For this purpose, we combine RNA-seq expression information and functional annotation on the protein level. All available protein-coding transcripts are annotated on the protein level in terms of functional characteristics (e.g. active sites, protein-protein interaction regions and domains). This annotation is based on existing knowledge of one of the proteins of a gene (SwissProt canonical protein isoform) and on evolutionary information.Combining the tissue RNA-seq data with the annotations, we identify cases where the highest expressed isoform is not the canonical isoform and try to characterise how the functional characteristics behave for this set of proteins. This is done via a custom functional conservation score.One of the key findings of this work is the observation of a bimodal distribution of the functional characteristics. That is, there is a tendency for alternative splicing to prefer either inclusion or exclusion of a functional characteristic in contrast to partial inclusion.	Proteins poster	Fundamental
P_Pr019	620	Fabian Sievers and Des Higgins	Fabian Sievers	Benchmarking Multiple Protein Sequence Alignments and the Effect of Guide-Tree Topology	Background: Multiple Sequence Alignments (MSAs) of large numbersof sequences are used in many bioinformatics analyses. However, thequality of progressive MSAs scales badly with the number of sequences.Methods: We show how the quality of MSAs decreases with largenumbers of sequences by benchmarking the quality of the alignment ofembedded reference sequences. One shortcoming of this benchmark isthat only a small fraction of sequences contributes to the qualityassessment of the MSA. We therefore present two schemes that eitheruse contact-map or secondary structure predictions based on the MSA asa measure of quality. These methods use all sequences and columns inthe alignment and are independent of potentially incorrectly curatedreference alignments.Results: The quality of MSAs decreases markedly for all alignersin our study, as the number of sequences is increased beyond a fewhundred sequences. Iteration can increase the useful range ofsequences to something like 1,000 sequences. Using high-quality HMMsseems to have the greatest effect in maintaining MSA quality at thisstage. The usefulness of chained guide-trees and high-qualitybackground HMMs can also be confirmed using our contact-map andsecondary structure prediction methods, which broadly correlate withscores derived from embedded crystal-based reference alignments.References: Fox G, Sievers F, Higgins DG (2015) Using de novo protein structure predictions to measure the quality of very large multiple sequence alignments. Bioinformatics; doi: 10.1093/bioinformatics/btv592	Proteins poster	Fundamental
P_Pr020	382	Po-Chia Chen and Jochen Hub	Po-Chia Chen	Biomolecular structure and dynamics via combined solution scattering experiments & atomistic simulations	X-ray and neutron solution scattering are powerful techniques that are capable of probing the solution behaviour of biomolecules. The measured scattering intensities contain information about the structural ensemble, both in terms of average structure and diversity. However, this information is disguised behind a global average over all conformations and orientations. Thus, measured SAS and WAXS patterns must be interpreted, ideally with independent atomic-level information to alleviate intrinsic ambiguity issues.We previously implemented an explicit-solvent approach in GROMACS to predict the ensemble SWAXS pattern of a biomolecule using molecular dynamics, and demonstrated the necessity of sampling at least picosecond and nanosecond-level freedoms in order to accurately reproduce experiment. Level of sampling depend on the underlying flexbility: from stable folds where accuracy is limited by sidechain and solvent sampling, up to intrinsically-disordered proteins, where accuracy is limited by the sampling of overall conformations. The MD integration also permits the use of SAXS data as constraints, which enables the direct isolation of structures consistent with a target SAXS pattern using related atomic coordinates as the starting conformation.A summary of above functionalities will be presented, along with planned extensions to integrate SANS techniques with contrast variation. We also plan to make capabilities available to integrative modelling workflows on HPC and cloud centers.	Proteins poster	Fundamental
P_Pr021	460	Gergely Gyimesi, Péter Závodszky and András Szilágyi	András Szilágyi	Calculation of configurational entropy differences from conformational ensembles using Gaussian mixtures	The configurational entropy of a molecular system is an important component of free energy that is often neglected in free energy calculations because of the inherent difficulty of the entropy calculation. The commonly used quasiharmonic method is unable to account for multiple basins and anharmonicities in the energy landscape. Here, we present a novel, conceptually simple approach to calculate the configurational entropy difference between two conformational ensembles (typically generated by molecular dynamics or Monte Carlo simulations) of a molecular system. The method estimates the probability density function of the system by a Gaussian mixture, using an efficient greedy learning algorithm along with a cross-validation based stopping criterion. Evaluating the method on conformational ensembles corresponding to substates of five small peptide systems, we found excellent agreement with the exact entropy differences obtained from a full enumeration of conformations. Compared with the quasiharmonic method and two other, more recently developed methods, the Gaussian mixture method yields more accurate results at smaller sample sizes. We illustrate the power of the method by calculating the backbone torsion angle entropy difference between disulfide-bonded and non-disulfide-bonded states of tachyplesin, a 17-residue antimicrobial peptide, and between two substates in the native ensemble of the 58-residue bovine pancreatic trypsin inhibitor. The Gaussian mixture method is a powerful and accurate approach for calculating configurational entropy differences for systems with complex energy landscapes. The program is written in Python and is available from the authors upon request.	Proteins poster	Fundamental
P_Pr025	322	Waqar Ali, Anatol Wegner, Robert Gaunt, Charlotte Deane and Gesine Reinert	Charlotte Deane	Comparison of large networks with sub-sampling strategies	Networks are routinely used to represent large data sets, making the comparison of networks a tantalizing research question in many areas. Techniques for such analysis vary from simply comparing network summary statistics to sophisticated but computationally expensive alignment-based approaches. Most existing methods either do not generalize well to different types of networks or do not provide a quantitative similarity score between networks.In contrast, alignment-free topology based network similarity scores empower us to analyse large sets of networks containing different types and sizes of data. Netdis is such a score that defines network similarity through the counts of small sub-graphs in the local neighbourhood of all nodes.Here, we introduce a sub-sampling procedure based on neighbourhoods which links naturally with the framework of network comparisons through local neighbourhood comparisons. Our theoretical arguments justify basing the Netdis statistic on a sample of similar-sized neighbourhoods. Our tests on empirical and synthetic datasets indicate that often only 10% of the neighbourhoods of a network suffice for optimal performance, leading to a drastic reduction in computational requirements. The sampling procedure is applicable even when only a small sample of the network is known, and thus provides a novel tool for network comparison of very large and potentially incomplete datasets.	Proteins poster	Fundamental
P_Pr026	365	R. Charbel Maroun, Pa Curmi and H El Shanti	R. Charbel Maroun	Consanguinity, genetic disease and molecular simulations	Two siblings born to a consanguineous couple with a previously un-described syndrome were identified. CLDN10 on chromosome 13 stood out as the best candidate gene. Re-sequencing of the coding region of CLDN10 and the flanking splice sites revealed a missense variation c.392C>T (NM_006984); p.S131L in claudin-10b, one of the alternatively spliced isoforms.The claudins are integral membrane proteins involved in the formation of the Tight Junction, which serves as a physical barrier to prevent solutes and water from passing freely through the paracellular space.To provide the molecular basis for this syndrome, we generated 3D models of claudin-10b, a 4-helix bundle. The direct effects of the p.S131L mutation in claudin-10b are a structural destabilization of the 4-helix bundle. In the cell, this should translate in the retention of the newly synthesized protein, given its inability to fold. Addressing of the protein to the plasma membrane should thus be impaired.This prediction was verified experimentally: the WT protein was observed at the plasma membrane after transfection and the label appeared stronger when two adjacent cells were transfected. On the contrary, when cells were transfected with the claudin-10b mutant, the plasma membrane was not labeled and the intercellular space appeared without any fluorescence.	Proteins poster	Fundamental
P_Pr027	424	Olga Zanegina, Evgeniy Aksianov, Andrei Alexeevski, Anna Karyagina and Sergei Spirin	Olga Zanegina	Conserved DNA-protein contacts formed by TATA-box binding proteins	TATA-box binding proteins (TBPs) are components of multiprotein complexes known as TFIID1. These complexes take part in transcription initiation of many genes of Archaea and Eukaryota. TBPs are two-domain proteins, 179–187 amino acid residues in length. In transcription initiation, TBPs specifically bind promoter regions containing sequences 5'-TATAWAWN-3', known as TATA-boxes.At the time, 34 structures of TBP from seven organisms are solved. Among them 25 are in complexes with DNA and 29 contain water molecules. Contacts of TBP with the DNA are highly conserved both in different structures of the same protein and in complexes of proteins from different organisms. We found 22 amino acid residues that form conserved hydrogen bonds and hydrophobic clusters at the DNA–protein interface. We also investigated conserved water molecules, both on DNA–protein interface and on the surface of the unbounded protein. We annotated a functional role of residues participating in the recognition of the DNA. The analysis was performed using services of the database of structures of DNA–protein and RNA–protein complexes NPIDB (http://npidb.belozersky.msu.ru/). The contact distribution on the DNA–TBP interface is in correspondence with the quasi-symmetry of N- and C-terminal domains of the protein. Although most contacts are symmetric, the N-terminal domain is connected with the DNA more tightly forming additional direct and water-mediated hydrogen bonds.As an example of possible applications of our results we predicted possible contacts with DNA of a homologous protein, TBP-L1, whose 3D structure is not available at the moment.	Proteins poster	Fundamental
P_Pr028	851	Václav Mareška and Vojtěch Spiwok	Václav Mareška	Development of the new pharmacophore model: test screening of inhibitors for COX-2 and KAT II	Using pharmacophores becomes an increasingly popular for the searching of new drugs. In comparison with traditional methods, pharmacophore models allow to be fast and efficient tool for virtual screening of large compound databases. Generally, pharmacophore models quantitatively characterize compounds by transformation of their structural characteristics into collective variables. This creates molecule "fingerprints" that can be easily compared. We have tried to design and implement the new pharmacophore model based on: CATS, SQUID and LIQUID models. We have used this model for screening of over 39 million compounds from ZINC database. Nowadays, we test the model for finding of new cyclooxygenase-2 (COX-2) and kynurenine aminotransferase II (KAT II) inhibitors with the same or even better biological activity compared to already known inhibitors. Together with docking calculations, we will test the pharmacophore model for screening another databases, searching new active molecules and will try to improve performance or efficiency of the model.	Proteins poster	Health
P_Pr029	601	Kenji Etchuya and Yuri Mukai	Kenji Etchuya	Environment Factor Depending on Each Sugar Type around O-glycosylation Sites in Mammalian Proteins	Glycosylation is a major post-translational modification and is important for protein folding, function, and enzyme activity. In O-glycosylation, motif residues (usually Ser or Thr) are modified by various kinds of sugars due to each glycosyltransferase in the Golgi body. The resulting sugars each promote a specific biological function and play different roles in living cells. Analysis of each sugar type will enable correlations between sugar type and biological function to be clarified. However, the characterization of the protein's primary sequences around each sugar type was weak and lacked consistency. An analysis of the environmental factors surrounding each sugar type is necessary to clarify the interaction between the glycosyltransferase and glycoprotein. Therefore, the environmental factors, composed of amino acids, were analyzed in this study.The sequence and structural data from mammalian proteins that undergo O-glycosylation was extracted from the Uniprot KB/Swiss-Prot 2015_03 and the Protein Data Bank (PDB) release 2015_03, respectively. The physicochemical environment constructed by amino acids around the O-glycosylation sites was investigated by analyzing the amino acid propensities depending on each sugar type within a unit ball in which center was an O-glycosylation site. The propensity of the amino acids was calculated and compared between each sugar types. Significant aromatic residues were found around each sugar, and the correlation between aromatic residues and sugar chains was analyzed. The environmental factors for each sugar type were discussed in this study.	Proteins poster	Fundamental

P_Pr030	355	Kliment Olechnovic and Ceslovas Venclovas	Kliment Olechnovic	Estimation of protein structure quality using contact areas derived from the Voronoi tessellation of atomic balls	In the absence of experimentally determined protein structure many biological questions can be addressed using computational structural models. However, the utility of protein structural models depends on their quality. Therefore, the estimation of both global quality and the quality of local regions of predicted structures is an important and as yet unsolved problem. One of the popular approaches to this problem is the use of knowledge--based statistical potentials. Such methods typically rely on the statistics of distances and angles of residue--residue or atom--atom interactions collected from experimentally determined structures. We present VoroMQA ("Voronoi diagram--based Quality Assessment"), a new method for the estimation of protein structure quality. Our method combines the idea of statistical potentials with the advanced use of the Voronoi tessellation of atomic balls. The new method uses contact areas instead of distances for describing and seamlessly integrating both explicit interactions between protein atoms and implicit interactions of protein atoms with solvent. In addition, VoroMQA utilizes the Voronoi tessellation of balls to describe the orientation of contacts. The method produces scores at atomic, residue and global levels, all in the fixed range from 0 to 1. Also, due to its design, our method evaluates structures of protein complexes as efficiently as monomeric structures. The latest version of VoroMQA was tested on the CASP11 data: the results showed that our method generally performs better than the other available methods using knowledge--based statistical potentials. The software implementation of VoroMQA is freely available as a standalone application and as a web--server.	Proteins poster	Fundamental
P_Pr031	597	Maciej Pajak, Clive R. Bramham and T. Ian Simpson	Maciej Pajak	Exploring spatio-temporal landscape of post-synaptic proteome diversification and functionalisation	Evolution of the post-synaptic proteome (PSP) can be traced back to primitive organisms that lack nervous systems and is thought to be responsible for the emergence of finely-tuned neural system function and behaviour in complex organisms, however these studies have only assessed evolution at the whole protein level. We present an evolutionary analysis of 1461 proteins, the complete human PSP as identified experimentally by Bayes et al. (2011). We focus on selected protein families and complexes, but also analyse the entire set of proteins in search for general patterns of selection spanning multiple subsets of post-synaptic proteins. Our custom analysis framework uses an integrative approach to study selection pressure, aggregating information inferred from models of branch, site, and branch-site selection which allow detection of previously overlooked signals of active diversifying pressure.Firstly, we evaluate the spatial distribution of selection pressure at single amino acid resolution and interpret these in relation to the location of functional domains and post-translational modification sites uncovering domain-level signatures of diversification and revealing strong candidates for downstream functional studies. Secondly, we use bootstrap clustering of PSP elements by their branch-by-branch selection pressure profiles to identify with high confidence distinct temporal patterns of episodic diversification shared by groups of proteins. We map these back to key divergence points in the tree of life allowing a detailed explanation of the rapid development of complex neural function in organisms such as primates, complementing and extending earlier hypotheses.	Proteins poster	Fundamental
P_Pr033	509	Eugenia Polverini, Ilaria Menozzi and Rodolfo Berni	Eugenia Polverini	HIGH STRUCTURAL AND FUNCTIONAL CONSERVATION BUT DIFFERENT LIGAND UPTAKE: THE ROLE OF THE HYDROPATHY PROFILE OF THE PROTEIN SURFACE	Cellular Retinol-binding Proteins (CRBP) type I and II are beta-barrel proteins that show very high structural conservation in spite of a moderately low sequence identity and a different tissue distribution. These retinol carriers play role in the maintenance of vitamin A homeostasis, but exhibit a different affinity for the ligand (100 folds higher for CRBP-I). However, the binding site of the two isoforms is highly conserved.The mechanism of ligand uptake was investigated by means of molecular dynamics simulations, initially positioning the ligand outside the protein. For both CRBPs, the portal region formed by alfa helix II and the two loops between CD and EF strands is involved in the uptake, with a partial unfolding of the helix II. Nevertheless, a different distribution of polar and hydrophobic residues clusters at the surface of the two proteins, in particular at the barrel lid made by helix I and II, favored two different entrance pathways. In CRBP I, the retinol enters the binding cavity through a hydrophobic passage between alpha helix II and CD and EF loops, while in CRBP II the ligand, driven by a few polar interactions, sinks in the hydrophobic region between the two alpha helices. Then, in both cases, several polar residues interacting with OH-group, attract the retinol deeply inside the binding cavity. Therefore, even if the retinol uptake involves the same region, that covers the binding pocket and is intrinsically flexible, the ligand finds the better entrance pathway according to the hydropathy features of the protein surface.	Proteins poster	Health
P_Pr034	797	Tamás Langó, Gergely Róna, Éva Hunyadi-Gulyás, Lilla Turiák, Julia Varga, László Dobson, Nóra Kucsma, György Váradi, János Molnár, László Drahos, Beáta G. Vértessy, Katalin F. Medzihradszky, Gergely Szakács and Gábor E. Tusnády	Tamás Langó	High througput experimental method to improve topology prediction of transmembrane proteins	AbstractTransmembrane proteins play a crucial role in signaling, ion transport, nutrient uptake, as well as in maintaining the dynamic equilibrium between the internal and external environment of cells. Despite their important biological functions and abundance, less than 2% of all determined structures are transmembrane proteins. Given the persisting technical difficulties associated with high resolution structure determination of transmembrane proteins, additional methods, including computational and experimental techniques remain vital in promoting our understanding of their structures, functions and interactions. The topology of transmembrane proteins defines the sequential position and orientation of transmembrane segments and the loops connecting them relative to the inner or outer sides of the membrane. The accuracy and reliability of in silico topology prediction algorithms can be significantly improved by incorporating experimental data as constraints. Therefore, generating such topology data could expedite structural modeling of transmembrane proteins. Here we report a novel, highly optimized high-throughput method for the generation of reliable experimental topology data for transmembrane proteins. Identification of covalently labeled cell surface amino acids by LC/MS/MS allowed the identification of extracellularly located protein segments, which were implemented in an improved computational method to provide accurate and reliable topology models for hundreds of human transmembrane proteins.ReferencesDobson, L., Reményi, I., and Tusnády, G. E. (2015) The human transmembrane proteome. Biol. Direct 10, 31	Proteins poster	Fundamental
P_Pr035	855	Zoran Sucur and Vojtech Spiwok	Zoran Sucur	Homology Modeling and Funnel Metadynamics in the study of oxytocin binding to its GPCR receptor	After being released from neurohypophyseal neurons, in the target tissues oxytocin binds to its GPCR receptor, which has not been studied in detail, yet. G-protein-coupled receptors (GPCRs) belong to very diverse and numerous receptor family, and are involved in vital cell signaling pathways. Different GPCR templates were used for homology modeling, and the best results were obtained for models based on orexin and adenosin a2a receptors. Using Schrodinger software, multiple stable conformations of oxytocin have been identified. In addition, we performed the docking of this hormone to the GPCR model receptor. Further studies of the oxytocin binding modes and its conformational changes upon binding to receptor were performed using Funnel metadynamics, which has proved to be a good technique used for enhancing the exploration of the ligands target binding site.The project was supported by Ministry of Education, Youth and Sports (COST action GLISTEN, CM1207, LD14133, Specific University Research MSMTNo. 20/2014, 21/2014 and 20/2015) and Czech Science Foundation (15-17269S). Computational resources were provided by the MetaCentrum under the program LM2010005 and the CERIT-SC under the program CentreCERIT Scientific Cloud, part of the Operational Program Research and Development for Innovations, Reg. no. CZ.1.05/3.2.00/08.0144.	Proteins poster	Health
P_Pr036	571	Tatsuki Kikegawa, Hiromu Sugita, Ryohei Nambu, Noritaka Kato and Yuri Mukai	Tatsuki Kikegawa	Identification of the subcellular localization factors of transmembrane proteins	Transmembrane proteins are typical internal membrane proteins spanning biomembranes including the endoplasmic reticulum (ER), Golgi, and plasma membranes. Their functions are essential to maintain homeostasis via signal transduction, membrane transport, and energy production. Their transmembrane regions usually consist of 10–30 hydrophobic amino acids, which are known as ER-targeting signals called signal-anchors. However, the mechanisms of transmembrane protein localization from ER to other organelles have not been elucidated. Understanding the mechanisms of protein subcellular localization is believed to be crucial for treatment of the incurable diseases resulting from erroneous subcellular localization.In this study, the amino acid propensity around signal-anchors was calculated to elucidate subcellular localization mechanisms of single-pass transmembrane proteins. The transmembrane protein dataset was classified into four groups: plasma membrane proteins, ER membrane proteins, Golgi membrane proteins, and proteins containing KDEL (KKXX) ER retention motif. The results of this analysis suggested that the amino acid propensity was related to the localization mechanisms because a remarkable bias of amino acid propensities was found in each group. These results were applied for predicting protein subcellular localization. The discrimination parameters of each group were evaluated by artificial GFP-signal-anchor fusion proteins. The GFP fusion proteins were expressed in HeLa cells, and the subcellular localization of these proteins was observed by a confocal laser fluorescence microscope.	Proteins poster	Fundamental
P_Pr037	697	Tomas Bastys, Vytautas Gapsys, Nadezhda Doncheva, Hauke Walter, Rolf Kaiser, Mario Albrecht, Bert Groot and Olga Kalinina	Tomas Bastys	Impact of point mutations on inhibitor affinity in HIV-1 protease	HIV (human immunodeficiency virus) protease is one of major targets of antiretrovital therapy, targeted by protease inhibitors (PIs). Through point mutations in protein sequence, a virus population acquires resistance to drugs. Effect of mutation on drug binding can be described in terms of change in drug binding free energy (ΔΔG) or change of the protein half maximal inhibitory concentration, also called resistance factor (RF). Predicting effect of a specific mutation on drug binding is essential for optimizing patient therapy. And understanding the specific mechanisms that influence affinity of the protein towards a PI is of important for development of novel drugs.In this work, we analysed a set of different combinations of known major resistance-associated mutations in HIV protease in complex with different PIs, for which experimental ΔΔG or RF measurements were available. For each combination, molecular dynamics simulations were used to calculate ΔΔG using Bennett'sAcceptance Ratio method. For a dataset of ten complexes we achieved a correlation coefficient of 0.81 between the theoretical and experimental values of ΔΔG. On a different dataset of eight protease-PI complexes where only RF measurements were available, combining information from ΔΔG calculations for complexes with the same mutations but different PIs, we were able to estimate values of RF, which were in most cases not significantly different from experimental measurements.Partial least-squares regression on molecular dynamics produced predictive models that were able to distinguish dynamics of wildtype and resistant proteases. These models illuminate different mechanisms that contribute to resistance against the PIs.	Proteins poster	Fundamental
P_Pr038	689	Maarten Reijnders, Vitor Martins Dos Santos and Peter Schaap	Maarten Reijnders	Improving functional annotation of microalgal proteins	Microalgae are promising organisms for the production of biobased compounds. However, to make the industrial production of these compounds competitive, we need to understand and improve the metabolic capabilities of microalgae [1]. The first step in understanding is a functional annotation of the proteins encoded in the genome. For a novel species, sequence similarity with proteins of known function from phylogenetic close-by model species is often used to transfer function. However, in absence of well-annotated close-by model species this is not a reliable way of assigning protein functions to microalgae.To reliably assign functions to microalgal proteins we have to go beyond the standard methods. We have designed a pipeline that utilizes multiple existing methods. Individual predictions are combined and compared by scoring for semantic similarities between the gene ontology terms obtained, followed by a machine learning algorithm over all the scores retrieved in the process.The performance of the pipeline on a test set of enzymes showed an improved true positive – false positive ratio compared to existing methods [2]. Additionally, compared to the same existing methods more proteins were annotated, with more annotations per protein.An additional benefit of this method is the modularity. In theory, other function prediction method can be incorporated in the processes of this pipeline. This allows for continuous improvement of annotation performance.1. M.J.M.F. Reijnders, "Green genes: bioinformatics and systems-biology innovations drive algal biotechnology". Trends in Biotechnology 32.12:617-626, 2-14.2. M.J.M.F. Reijnders, "Algal omics: The functional annotation challenge". Current Biotechnology 4.4:457-463, 2015.	Proteins poster	Fundamental
P_Pr039	760	Eda Suku, Mattia Di Giacobbe, Behnoosh Bahadori, Stefano Capaldi, Mario R. Buffelli and Alejandro Giorgetti	Eda Suku	In silico deorphanization of the GPR3 receptor	Introduction: Alzheimer's Disease is a neurodegenerative disease (ND), characterized by loss of brain connectivity, memory and cognitive functions. Recently, G-protein coupled receptor 3 (GPR3) was identified as regulator of Aβ plaques through the β-arrestin 2 pathway1. GPR3 is an orphan receptor and a deep investigation of its function is still missing. Here we present the identification of two putative GPR3 endogenous ligands and structural insights into the binding pocket using state of the art techniques.Methods: Homology modeling and docking were carried out through the GOMoDo web-server2. The programs OMEGA3 and ROCKS3 were used to perform chemoinformatics searches on ZINC and Human Metabolome Databases.Results: GPR3 model was validated against experimental data on non-endogenous ligands: DPI4 and AF6439445. These molecules were used as starting compounds for chemoinformatics studies. Two endogenous ligands, i.e. beta-carboline and 1-methyladenine, present in different brain pathways and involved in neuronal damage, have been identified. Docking studies of these ligands allowed us to characterize residues putatively involved in receptor-ligand interaction.Conclusions: We aimed to in silico deorphanize GPR3 and characterize its binding cavity. We identified two endogenous ligands and several putative critical residues. Further investigations and experimental informations are being carried out to validate the results and to better characterize the GPR3 function.References:1Thathiah A. et al. Nature medicine (2013): 43-492Sandal M. et al. PLoS One (2013): e740923http://www.eyesopen.com/4Ye C. et al. Journal of Pharmacology and Experimental Therapeutics (2014): 437-4435Jensen T. et al. Bioorganic & medicinal chemistry letters (2014): 5195-5198	Proteins poster	Biotechnology
P_Pr041	628	Dinithi Sumanaweera and Dr. A. Shehan Perera	Dinithi Sumanaweera	In silico prediction of protein function for Saccharomyces Cerevisiae; an ensemble approach	Protein function annotation is vital for identifying disease causative factors and for solving mysteries behind biological system complexities. As manual annotation requires costly and laborious in-vitro methods, in-silico protein function prediction is preferred nowadays. According to literature, one in five yeast mitochondrial proteins are known to be human disease related. We present a weighted heterogeneous data ensemble to classify Saccharomyces Cerevisiae proteins under 'Mitochondrial Organisation' in Gene Ontology (GO). It consists of five euclidean-distance based nearest neighbour models and three affinity-based neighborhood models; utilizing protein properties data, four gene expression datasets and physical/genetic interactions. 239 current GO annotations and 3887 gold standard negative annotations from literature were used to train the base learners. The overall prediction is the weighted average of posterior probabilities outputted by the base models. The weights are determined by a Genetic algorithm (GA) for obtaining the optimal AUC value under ROC. All evaluations were performed using leave-one-out cross-validation for 10 samples; each containing all positive proteins and a random negative protein sample, with 1:1 class ratio. The optimal k parameter for nearest neighbour models was decided as 22 upon empirical results obtained by varying k from 1 to 25. The base models show a substantial level of disagreement with a mean Fleiss Kappa statistic of ~0.2816. The GA-weighted ensemble gives ~14.3% (from ~78.34% to ~89.57%) improvement of the best performing base classifier, whereas only ~13.1% improvement can be seen with an equal-weighted ensemble.	Proteins poster	Fundamental
P_Pr042	533	Fabrizio Pucci, Raphaël Bourgeas, Jean Marc Kwasigroch and Marianne Rooman	Fabrizio Pucci	In-silico prediction of protein thermal stability changes upon point mutations using HoTMuSiC	IntroductionThe ability to rationally modify proteins in order to increase their thermal stability is one of the main goals of protein design, which has interesting applications in a wide series of biomedical and biotechnological processes.We present a newly developed bioinformatics tool that, using as input the three-dimensional (3D) structure of the protein and, when available, its melting temperature (Tm), is able to predict rapidly and accurately the impact of amino acid substitutions on this temperature.MethodsThe key ingredients of our methodology are statistical potentials that are knowledge-driven mean force potentials (PMF) extracted from a dataset of experimentally resolved 3D protein structures. They are linearly combined using an artificial neural network (ANN) with sigmoid activation functions that depend on the solvent accessibility of the mutated residues. If the melting temperature of the protein is known, we use in addition temperature-dependent statistical PMFs that reflect the (melting)-temperature dependence of the amino acid interactions. They are combined using a triple-layer ANN, in which the activation functions of the first layer depend on the solvent accessibility, while those for the second layer are parabolic functions of the protein's number of residues and melting temperature.ResultsThe performance of our method is evaluated in 5-fold cross validation on a dataset of 1626 mutations and yields a root mean square deviation between predicted and experimental ΔTm's of about 4°C. The addition of evolutionary information to the model and the analysis of the relations between thermal and thermodynamic stability changes are also carefully discussed.	Proteins poster	Biotechnology Fundamental
P_Pr044	670	Nesrine Chakroun, Cheng Zhang and Paul Dalby	Nesrine Chakroun	Insights into the intrinsic Stability of a Therapeutic Fragment Antibody by Molecular Dynamics Simulations	Biopharmaceuticals or therapeutically relevant proteins have become one of the fastest growing parts of the pharmaceutical industry. These innovative molecules are more complex than conventional drugs and their processing is much more demanding.The analytical characterization of these new drugs is a fundamental step in the early prediction of their behavior in bioprocesses. This research project aims to develop a framework to improve candidate design and selection at early stages of development by establishing a set of critical analysis and identifying key properties (intrinsic and extrinsic) allowing the prediction of candidates behaviour in large-scale bioprocesses.Our multidisciplinary approach combines the computational analysis (sequence analysis, Molecular Dynamics simulations and docking) and the biophysical characterization of a set of Fragment antibody (Fab) mutants. In particular MD simulations were used to investigate the effects of pH, temperature and mutations in the stability of Fab. This allowed the identification of several key regions and residues in the stability of the molecule which were targeted experimentally to enhance candidate's stability. The effect of formulation was also investigated highlighting the role of electrostatics and salt bridges in Fab stability and folding. Additionally, aggregation kinetics studies were carried out at a wide range of temperature, pH and ionic strength allowing the determination of a model for Fab aggregation	Proteins poster	Health
P_Pr045	849	Gift Nuka, Simon Potter, Siew-Yit Yong, Maxim Scheremetjew, Alex Mitchell, Matthew Fraser and Rob Finn	Gift Nuka	InterProScan 5: Large scale protein function classification	InterPro (http://www.ebi.ac.uk/interpro/) is a freely available resource that is used to classify sequences into protein families and to predict the presence of important domains and sites. InterProScan (https://www.ebi.ac.uk/interpro/interproscan.html) is the underlying software application that allows both protein and nucleic acid sequences to be scanned against InterPro's predictive models (signatures), which are provided by the resource's member databases.Recently, both the Conserved Domain Database (CDD) and Structure-Function Linkage Database (SFLD) have joined InterPro as new member databases. InterProScan has been updated accordingly, incorporating CDD's curated models that use position specific scoring matrices (PSSMs) to represent protein domains, which tend to be more functionally specific than some of the models already used in InterPro. SFLD's hidden Markov models that offer structure-function mapping have also been incorporated. SFLD models allow evolutionary classification of related enzymes according to shared chemical functions to determine conserved active sites.Here, we present these recent developments and performance improvements to InterProScan.Optimisation in the pipeline filters and database query refinements have also resulted in improved throughput for large-scale protein sequence analysis and accelerated InterProScan domain searches by several orders of magnitude.	Proteins poster	Biotechnology Fundamental

P_Pr046	459	Sirawit Ittisoponpisan, Eman Alhuzimi, Michael Sternberg and Alessia David	Sirawit Ittisoponpisan	Landscape of pleiotropic proteins causing human disease: structural and system biology insights.	Pleiotropy is the phenomenon by which the same gene can result in multiple phenotypes. Pleiotropic proteins are emerging as important contributors to both rare and common disorders. Despite this, little is known on the pathogenetic mechanisms underlying pleiotropy and the characteristic of pleiotropic proteins.We analysed disease-causing proteins reported in Uniprot and observed that 12% are pleiotropic (mutations in the same protein cause more than one disease). Pleiotropic proteins were more likely to be essential and have a higher number of interacting partners compared to non-pleiotropic proteins. Moreover, significantly more pleiotropic than non-pleiotropic proteins contained at least one intrinsically long disordered region of over 50 residues in length (p<0.001). Pleiotropic proteins were enriched in deleterious mutations and rare polymorphisms, but not in common polymorphisms. Deleterious mutations occurred mainly in structurally ordered regions. Deleterious mutations occurring in structurally disordered regions were more commonly found in pleiotropic, rather than non-pleiotropic proteins. Finally, we observed that proteins involved in the pathogenesis of neoplasms, neurological and circulatory diseases, and congenital malformations were more likely to be pleiotropic, whereas proteins causing endocrine and metabolic disorders were more likely non-pleiotropic.In conclusion, this study suggests that pleiotropic proteins represent a biologically different class of proteins compared to non-pleiotropic proteins and are an important contributor to human disease. This study provides a better understanding of pleiotropic proteins and their genetic variants, which could greatly aid in the interpretation of genetic studies and drug design.	Proteins poster	Fundamental
P_Pr047	604	Chloé Dequeker, Raffaele Raucci, Elodie Laine and Alessandra Carbone	Chloé Dequeker	Large scale analysis of protein interactions	Protein-Protein Interactions (PPI) are at the heart of processes and their understanding is of utmost importance to facilitate drug design and characterize the mechanisms underlying certain diseases.In this context, our team works on the Help Cure Muscular Dystrophy (HCMD) project, whose aim is to uncover new pathways responsible for the muscular dystrophy by developing a discriminating power over the interacting and non interacting complexes. A complete cross-docking (CCD) has then been realised over 2200 proteins with the help of the World Community Grid (WCG), generating more than 900 billions conformations over 2.5 millions different complexes.In parallel of these computations, our team developed a new method JET² to predict interacting surfaces at large scale (Laine and Carbone, 2015), using different criteria based on residue conservation, physico-chemical properties and the geometrical aspect of the protein structure. JET² has been run over more than 20.000 different chains for which a PDB structure is available.We present new ways to link the two different problems of the prediction of protein interaction sites and the discrimination of interacting partners through optimisation of JET² prediction as well as using different scoring methods. Our work also sheds some light on interactions of proteins with multiple partners, which will be a principal factor in the analysis of the HCMD results.	Proteins poster	Biotechnology
P_Pr048	679	Nicholas Furnham, Natalie Dawson, Syed Rahman, Janet Thornton and Christine Orengo	Nicholas Furnham	Large-Scale Analysis Exploring Evolution of Catalytic Machineries and Mechanisms in Enzyme Superfamilies	Enzymes, as nature's catalysts, are crucial to life. How they have evolved to undertake their different chemical reactions is of great interest to a wide range of biological disciplines. Over 100 years of detailed biochemistry studies combined with the large volumes of sequence and protein structural data now available, means we are able to perform large-scale analyses to address this question. Using sophisticated tools relating sequences and structures across thousands of genomes though phylogenetic analysis and novel measures of functional similarity we have compiled information on all experimentally annotated changes in enzyme function within 379 structurally defined protein domain superfamilies, linking the changes observed in functions during evolution to changes in reaction chemistry. Using analysis of modifications in reaction chemistry and enzymes active sites we have observed that some superfamilies have changed the reactions they perform without changing catalytic machinery. In others large changes of enzyme function have been brought about by significant changes in catalytic machinery. Interestingly, in some superfamilies relatives perform similar functions but with different catalytic machineries. The collected data and analysis has been developed into a community resource (www.funtree.info). This analysis highlights characteristics of functional evolution across a wide range of superfamilies. It also provides insights that will be useful in predicting the function of uncharacterized sequences as well as the design of new synthetic enzymes.	Proteins poster	Fundamental
P_Pr049	499	Daniele Raimondi, Andrea Gazzo, Marianne Rooman, Tom Lenaerts and Wim Vranken	Daniele Raimondi	Multilevel biological characterization of exomic variants at the protein level significantly improves the identification of their deleterious effects	There are many predictors capable of identifying the likely phenotypic effects of single nucleotide variants (SNVs) or short in-frame Insertions or Deletions (INDELs) on the increasing amount of genome sequence data. Most of these predictors focus on SNVs and use a combination of features related to sequence conservation, biophysical, and/or structural properties to link the observed variant to either neutral or disease phenotype. Despite notable successes, the mapping between genetic variants and their phenotypic effects is riddled with levels of complexity that are not yet fully understood and that are often not taken into account in the predictions, despite their promise of significantly improving the prediction of deleterious mutants.We present DEOGEN, a novel variant effect predictor that can handle both missense SNVs and in-frame INDELs. By integrating information from different biological scales and mimicking the complex mixture of effects that lead from the variant to the phenotype, we obtain significant improvements in the variant-effect prediction results. Next to the typical variant-oriented features based on the evolutionary conservation of the mutated positions, we added a collection of protein-oriented features that are based on functional aspects of the gene affected. We cross-validated DEOGEN on 36 825 polymorphisms, 20 821 deleterious SNVs, and 1038 INDELs from SwissProt. The multilevel contextualization of each (variant, protein) pair in DEOGEN provides a 10% improvement of MCC with respect to current state-of-the-art tools. The software and the data presented is available at http://ibsquare.be/deogen.	Proteins poster	Health
P_Pr051	711	Rashmi Hazarika and Vera van Noort	Rashmi Hazarika	Network evolution of MADS-domain protein interaction network	In protein-protein interaction networks, the nodes symbolize interacting proteins while the edges relate to the physical interactions between these proteins. A gain of an edge between two nodes denotes the appearance of a new functionality while losing a subset of their initial interactions symbolizes functional divergence as when duplicate copies of a protein evolve to bind different interaction partners. In this study, we chose the MADS-domain transcription factors which rely on coiled coil interactions. The study of these proteins would help us understand plant evolution better, as proliferation of these proteins and successive diversification of protein functions may explain how modern day Angiosperms evolved. The ancestral nodes for the MADS-domain proteins were estimated, resurrected and their interactions experimentally verified before and after whole genome duplication. The Yeast-2-Hybrid system was used to define the protein-protein interactions of 9 resurrected ancestral MADS-box gene lineages (SEP3, SEP1/2/4, AP1, AP3, PI, AG, STK, SVP and SOC1) before γ triplication. After the γ triplication event, some of these genes duplicated or triplicated, giving rise to a total of 17 post γ triplication MADS-box genes. Using random networks, the various events of Whole Genome Duplications (WGDs) and network dynamics were simulated on an evolutionary time scale using age estimates from literature and the empirical probabilities measured from the actual Y2H networks. The results were directly compared to extant networks of Arabidopsis thaliana and Solanum lycopersicum. We observed a scale free and a highly modular network topology in the simulated networks.	Proteins poster	Fundamental
P_Pr052	775	François Ancien, Maxime Godfroid, Georges Coppin, Fabrizio Pucci and Marianne Rooman	François Ancien	Neural network-based predictions of deleterious human variants derived from protein structures and free energy estimations	Many predictors have been developed to predict the deleteriousness of mutations in the human exome, often exclusively based on the protein sequences and their evolutionary features. However, the explanatory power of these methods in terms of the physical effect that the mutations have on the molecular phenotype is usually quite limited – although such insight is a prerequisite for the development of personalized treatments. Here we analyzed what relevant information the protein structure and stability can add in this context. For that purpose we used a dataset of human variants that are annotated as deleterious or neutral in proteins for which the 3-dimensional structure is available. In a first step we estimated the thermodynamic and thermal stability changes caused by the mutations, using the PoPMuSiC and HoTMuSiC programs, which use artificial neural networks (ANN) and linear combinations of statistical mean-force potentials. These stability changes upon mutations were shown to correlate significantly with the deleteriousness of the mutations: the more destabilizing, the more deleterious, with a balanced accuracy of about 0.6. In a second step, we built on PoPMuSiC and HoTMuSiC to develop a new predictor that focuses on deleteriousness prediction. We implemented different types of ANNs, and in particular probabilistic and echo state networks, in an attempt to catch all the complex information contained in the dataset and to improve the prediction performances. The highest scores, estimated in cross validation, are significantly higher than that of PoPMuSiC and HoTMuSiC and exceed 0.7. This performance is comparable to that of purely evolutionary-based methods, with however the advantage of a better understanding of the biophysical effects that cause the disease.	Proteins poster	Fundamental Health
P_Pr055	381	Olga Voitenko, Andi Dhroso, Anna Feldmann, Dmitry Korkin and Olga Kalinina	Olga Voitenko	Patterns of amino acids conservation in human and animal immunodeficiency viruses	Motivation: Due to their high genomic variability, RNA viruses and retroviruses present a unique opportunity for detailed study of molecular evolution. Lentiviruses, with HIV being a notable example, are one of the best studied viral groups: hundreds of thousands of sequences are available together with experimentally resolved three-dimensional structures for most viral proteins. In this work, we use these data to study specific patterns of evolution of the viral proteins, and their relationship to protein interactions and immunogenicity.Results: We identify extremely conserved and extremely variable clusters of amino acid residues on the surface of proteins from HIV and other animal immunodeficiency viruses. These clusters turn out to be located on the interaction interfaces of viral proteins with other proteins, nucleic acids or low molecular-weight ligands, both in the viral particle and between the virus and its host. In the immunodeficiency viruses, the interaction interfaces are not more conserved than the corresponding proteins on average, and we show that extremely conserved clusters coincide with protein-protein interaction hotspots, predicted as the residues with the largest energetic contribution to the interaction. Extremely variable clusters have been identified here for the first time. In the HIV-1 envelope protein gp120, they overlap with known antigenic sites. These antigenic sites also contain many residues from extremely conservedclusters, hence representing a unique interacting interface enriched both in extremely conserved and extremely variable clusters of residues. This observation may have important implication for antiretroviral vaccine development.	Proteins poster	Fundamental Health
P_Pr056	458	Rosalba Lepore, Agnieszka Obarska-Kosinska, Alfredo Iacoangeli and Anna Tramontano	Rosalba Lepore	PepComposer: computational design of peptides binding to a given protein surface	There is a wide interest in designing peptides able to bind to a specific region of a protein with the aim of interfering with a known interaction or as starting point for the design of inhibitors.Structure-based strategies usually consists in analysing the interacting region from a complex of the target protein with a protein or a peptide trying and identifying a contiguous 'peptide like' region of the partner to be used as starting point. However, if no complex structure is available, one has to recur to de novo design methods and therefore needs to select an appropriate backbone, optimize its relative orientation with respect to the target protein and its sequence (1).To simplify and streamline this process, we developed PepComposer, a computational pipeline for the design of protein-binding peptides that only requires the target protein structure and an approximate definition of the binding site as input. We first select appropriate backbones from monomeric proteins based on previous observations (1) and use a Monte Carlo procedure to design optimal sequences for the identified peptide scaffolds. Peptides are then selected according to the predicted binding energy.PepComposer is fully automatic, available as a web server (http://biocomputing.it/pepcomposer/webserver) and can effectively reproduce known protein-peptide interactions (2).1. Vanhee P., et al. Computational design of peptide ligands. Trends Biotechnol. 2011;29:231-239.2. Obarska-Kosinska A., et al. PepComposer: computational design of peptides binding to a given protein surface. Nucl. Acids Res. (2016) doi: 10.1093/nar/gkw366	Proteins poster	Fundamental
P_Pr057	636	Emilie Neveu, David Ritchie, Petr Popov and Sergei Grudinin	Emilie Neveu	PEPSI-Dock : A Detailed Data-Driven Protein-Protein Interaction Potential Accelerated By Polar Fourier Correlation	Docking prediction algorithms aim at finding the native conformation of a complex of proteins, knowing their unbound structures. Most of the existing predictions the results of a combination of sampling and scoring methods, adapted to different scales. Here we present PEPSI-Dock (Polynomial Expansion of Protein Structures and Interactions for Docking), which improves the first stage of the docking pipeline, being more accurate at the beginning of the docking process, which thus sharpen up the final predictions. Indeed, the method benefits from the precision of a very detailed data-driven model of the binding free energy used with a global and exhaustive rigid-body search space. While being accurate, our computations are among the fastest ones by virtue of the sparse representation of the pre-computed potentials and FFT-accelerated sampling technics. PEPSI-Dock runs in 5-20 minutes on a modern laptop and can be easily extended to other types of interactions.	Proteins poster	Health
P_Pr059	356	Thanh Binh Nguyen and M.S. Madhusudhan	Thanh Binh Nguyen	Prediction of polyproline type II helices receptors	Polyproline type II helices (PPII) are a less common secondary structure of proteins than α helix and β sheet. There is no internal backbone hydrogen bond interaction in this conformation. As a result, the carbonyl and amide groups along the PPII helices prefer to make intermolecular interaction. And hence, PPII mediates many protein-peptide or protein-protein interactions in signalling pathway, immune response, cell-cell communication. There is an abundance amount of proteins which are well-known to bind PPII, including MHC, SH3, WW, EVH1, profilin and GYF domains. These PPII bound proteins share geometry and biophysical features. Using the knowledge from the known PPII-bound families the aim of this study is to detect the PPII binding site in a query protein. This approach could help to identify a new PPII-bound protein.	Proteins poster	Fundamental
P_Pr060	648	Thach Nguyen and Michael Habeck	Thach Nguyen	Probabilistic model for segmentation of protein structures	Motivation: Large-scale conformational changes in proteins are implicated in many important biological functions. These structural transitions can often be rationalized in terms of relative movements of rigid domains. There is a need for objective and automated methods that identify rigid domains in sets ofprotein structures showing alternative conformational states.Results: We present a probabilistic model for detecting rigid-body movements in protein structures. Our model aims to approximate alternative conformational states by a few structural parts that are rigidyltransformed under the action of a rotation and a translation. By using Bayesian inference and Markov chain Monte Carlo sampling, we estimate all parameters of the model, including a segmentation of the protein into rigid domains, the structures of the domains themselves, and the rigid transformations thatgenerate the observed structures. We find that our Gibbs sampling algorithm can also estimate the optimal number of rigid domains with high efficiency and accuracy. We assess the power of our method on severalthousand entries of the DynDom database and discuss applications to various complex biomolecular systems.Availability:The Python source code for protein ensemble analysis is available at:https://github.com/thachnguyen/motion_detection/.	Proteins poster	Biotechnology Fundamental
P_Pr061	793	István Reményi, László Dobson and Gábor E. Tusnády	István Reményi	Profile modeling and multiple sequence alignment of transmembrane proteins	Transmembrane proteins are involved in energy production, signal transduction, cell-cell interaction, cell-cell communication. They are frequent targets for pharmaceuticals, therefore knowledge about their properties and structure is crucial. However, less than 2% of all determined protein structures belongs to transmembrane proteins, thus computational approaches have to be utilized for topology prediction and structure modelling.Analyzing a protein may begin with searching for homologous sequences, namely for entities with statistically significant similarity. There are several methods for homology detection, among which profile modeling exceeds in terms of capability of capturing highly conserved regions. As a result, more accurate alignments can be created from any unaligned set of sequence, and more thorough analysis can be performed. Hidden Markov Models have already been applied to homology search, but their training problem can be considered NP hard. In such a model, to handle the high number of variables, the different training approaches are either local optimization techniques which incorporated task-specific additional (e.g. structure) information, numerous tuning parameters, or just based on an otherwise determined multiple sequence alignment.Previously we have launched the Human Transmembrane Proteome database, which contains topology and structure information about the human α-helical transmembrane proteins. Our aim is to find a general optimization technique to build TM profile model(s) with, to represent the whole proteome to create a starting point for further investigations.The Human Transmembrane ProteomeLászló Dobson, István Reményi and Gábor E. Tusnády (2015)Biology Direct, 10:31	Proteins poster	Fundamental
P_Pr062	720	Diego Alonso-Martinez and Peter Dimaggio	Diego Alonso-Martinez	Profiling the methylome targets of histone lysine methyltransferases	Histone post-translational modifications (PTMs) are epigenetic marks critical in the regulation of gene expression that are regulated by various classes of enzymes including histone lysine methyltransferases (HKMTs). HKMTs catalyse the transfer of a methyl group from S-adenosyl methionine (SAM) to a specific histone lysine target. Due to their overlapping but non-redundant functions, there is current no way to decisively assess which HKMT is responsible for an observed methylation event. This lack of understanding prevents the development of more specific treatments for epigenetic, PTM-related diseases, such as cancer.In this work we propose to engineer the first cellular HKMT methylome profiling assay by combining the classical "bump and hole" approach with extensive bioinformatics and computational modelling of enzyme-cofactor candidate pairs prior to mass spectrometry (MS) validation. A multiple sequence alignment of all post-SET domain containing HKMT sequences alongside a detailed analysis of the crystallographic structure of G9a-SAH (PDB: 2O8J) identified a bulky residue in G9a that could be mutated to create a "hole" for a modified SAM cofactor. Protein folding simulations using Rosetta were used to assess structural validity of the mutant candidates, and corresponding SAM analogues with a "bump" to match the "hole" were designed based on the generated models.In vitro experiments utilising recombinant G9a with our engineered SAM analogues have demonstrated suitable cofactor selectivity against endogenous HKMTs, which supports the feasibility of this approach. This study highlights the importance of computational simulations in the development of more accurate assays to characterise the methylome targets of HKMTs.	Proteins poster	Health

P_Pr063	718	Alexander Smolyakov, Ilya Altukhov, Sergey Gavrilov, Ivan Butenko, Olga Pobeguts, Ilya Kublanov and Dmitry Alexeev	Alexander Smolyakov	Quantitative profiling of membrane-associated proteins in Meliioribacter roseus P3M-2	Meliioribacter roseus P3M-2 is recently discovered gram-negative bacteria characterized as a new species of Meliioribacteraceae family within the Ignavibacteriae phylum. The complete sequence of the M.roseus genome was recently released and showed presence of genes involved in adaptation to the extreme conditions.Currently proteomic studies widely use mass-spectrometry analysis methods. These methods are mostly applied for investigating protein-protein interactions and post-translational modifications, as well as organism proteome inventory; they also offer strategies for quantitative and qualitative proteomic and proteogenomic analysis.In this study cells of M.roseus P3M-2 were grown by aerobic respiration and maltose fermentation at strictly anaerobic conditions. Each culture was grown in three biological replicates, independently, to give a total of six specimens. We carried out an in-depth quantitative proteogenomic analysis of M.roseus P3M-2 based on shotgun LC-ESI-MS/MS data. In total 199,894 tandem mass spectra were obtained. The 1,127 proteins were identified by two and more peptides across all experiments. The quantitative proteome analysis revealed 239 significantly different membrane-associated proteins between cells grown at aerobic and anaerobic conditions.Proteins were classified according to the Gene Ontology annotations BP, CC and MF. 28 GO terms were significantly enriched. KEGG pathway analysis was performed to determine the functional interactions of differentially expressed proteins between aerobic and anaerobic conditions. Wilcoxon test were performed for directed and and undirected regulation. 4 maps were significantly enriched (p.value < 0.05).Using proteogenome approach we have defined 3 novel protein-coding genes, coordinates of 9 genes were reannotated and peptides located in 10 pseudogenes were identified.	Proteins poster	Fundamental
P_Pr064	728	Francesca Nadalin and Alessandra Carbone	Francesca Nadalin	Residue propensity and local geometry of the interface contacts define the specificity of protein-protein interactions	Obtaining structures of protein complexes experimentally requires a lot of effort. For this reason, reliable methods for modeling PPI in silico are envisaged. Protein docking experiments output a long list of possible conformations, thus properly scoring them is essential for further studies.Previous works showed the applicability of pair potentials to the scoring of docking decoys [Moal 2013].We define new pair potentials as the contribution of two terms: the one derived by the observed contact distribution at the interface, the other representing the likelihood of residues to be located at the interface [Negi 2007].Pair potentials are computed on 230 experimental structures [Vreven 2015] and we call our method CIPS, Combined Interface Propensity for decoy Scoring. CIPS is compared to other methods, respectively based on pair potentials [Glaser 2001, Pons 2013, Mezei 2015] and atomic potentials [Krissinel 2007,Pierce 2007]. CIPS turns out to outperform all tested methods on decoys sets obtained both with all-atom and coarse-grain rigid docking. Further improvement is observed when decoys scoring is done with a combination of CIPS and atomic potentials.Our method is fast, accurate, and robust upon decrease in the level of detail of protein structure representation. This allows important application inthe promising, but not yet deeply explored, field of large-scale cross docking.	Proteins poster	Fundamental
P_Pr065	421	Gabriele Orlando, Daniele Raimondi, Tom Lenaerts and Wim Vranken	Gabriele Orlando	RIGAPOLLO, A HMM-SVM BASED APPROACH TO SEQUENCE ALIGNMENT	Reliable protein alignments are a central problem for many bioinformatics tools, such as homology modeling. Over the years many different algorithms have been developed and different kinds of information have been used to align very divergent sequences. Here we present a pairwise alignment tool, called Rigapollo, based on pairwise HMM-SVM, which can include different types of information in the alignment process. The model is composed by 7 states: a M (match), and six G (gap) states, three for the first sequence and three for the second one. For each amino acid in the sequences, we define an N-dimensional feature vector to describe it. That vector can be defined using any kind of information, from evolutionary (i.e. PSSM) to dynamics predictions. While standard pairwise HMMs require the definition of a finite and discrete alphabet of observable states, our model works directly using these feature vectors (that can be both orthonormal or not orthonormal). We define the emission probability using a SVM trained to discriminate matches from miss-matches and gaps from non-gaps positions. We tested our algorithm on two benchmark datasets of very divergent proteins, one based on Balibase and the other based on Sabmark. Rigapollo improves the quality of the predicted alignments from 20 to 38% respect to the most used state of the art pairwise alignment tools.	Proteins poster	Fundamental
P_Pr066	488	Qingzhen Hou, Paul De Geest, Wim Vranken, Jaap Heringa and K. Anton Feenstra	Qingzhen Hou	Seeing the Trees through the Forest: Sequence-based Homo- and Heteromeric Protein-protein Interaction sites prediction using Random Forest	Motivation: To fulfil biological functions, proteins bind to their partners via specific amino acids.Investigation of the properties and sequential information of these residues is important to reveal the mechanisms of protein-protein interactions and protein functions. These properties, derived from the interacting amino acids at sequence level, are usually exploited as features for machine learning methods to predict protein interacting positions. In this paper, we include two novel features (backbone flexibility and Sequence Specificity) predicted from sequences for protein interface prediction and evaluate the importance of different features using Random Forest.Results: We observe that there is no single sequence feature which enables to pinpoint interacting sites. However, combination of different properties does help the interface prediction. After selecting and integrating multiple features, we developed a Random Forest predictor which is able to distinguish interface and other residues with AUC of ROC plot at 0.72 in our homomeric test-set, which is better than other sequence-based methods. Moreover, when applied to identify interfaces of an independent heteromeric dataset, our method performs slightly better than the best sequence-only predictor. Thus, our predictor trained on homodimeric proteins can not only predict homodimeric interfaces, but is also able to locate interface residues in the heterodimers which suggested that our predictor captures the common properties of both homodimer and heterodimer interfaces.	Proteins poster	Fundamental
P_Pr067	521	Wim Vranken, Daniele Raimondi, Gabriele Orlando and Rita Pancsa	Wim Vranken	Sequence-based prediction of protein early folding residues	We present EFoldMine, a novel protein sequence-based predictor of early folding regions based on the Start2Fold database and the DynaMine predictions of protein backbone rigidity. EFoldMine reaches an AUC of 0.808 for detecting early folding residues, over a 27-fold set of 30 proteins. We observe that first, amino acids involved in amyloid formation have a higher tendency to fold early according to our predictions. Second, there is a weak correlation with folding speed, especially for two-state folders. Third, the predictions especially pick up residues that form extensive contacts in the folded conformation of the protein, less so than residues that become buried. Finally, residues with high covariance signals in the PSICOV contact prediction dataset tend to be in predicted early folding regions. On a proteome scale, the incidence of predicted early folding regions decreases with protein length for a set of human protein domains from PFAM. Overall, our sequence-based early folding prediction provides a novel picture of the residues in the unfolded protein that are inclined to form stable structural elements purely based on local sequence interactions. This view of the statistical behavior of proteins, prior to the formation of highly specific defined interatomic contacts in the folded protein, allows the incorporation of a different kind of information in structural bioinformatics approaches, which are currently mostly based on folded protein structures, and should stimulate further advances in the field.	Proteins poster	Fundamental
P_Pr068	787	Miguel Correa Marrero, Richard G.H. Immink, Dick de Ridder and Aalt D.J. van Dijk	Miguel Correa Marrero	Simultaneous prediction of protein-protein contacts and interaction partners	Protein-protein interactions underlie virtually any biological process. How proteins interact with each other is therefore a fundamental question in biology. However, techniques that give fine-grained information about protein-protein interactions are low-throughput and labour-intensive, which makes the development of in silico approaches attractive.One way to approach the problem is to exploit the phenomenon of coevolution. Protein-protein interaction leads to the coevolution of the interfaces between the interaction partners, meaning that there are correlations between their sequences. From these correlations, one can deduce which residues are involved in the interaction interfaces. This can be done by applying statistical models to multiple sequence alignments of homologs of the proteins of interest.However, one can easily introduce pairs of sequences that have lost the interaction, or paralogs. This introduces noise in the analysis and has limited the application of these coevolutionary approaches. To surpass this obstacle, we are developing a novel approach.Our approach combines traditional correlated mutation analysis with the expectation-maximization algorithm. For each sequence pair in the input alignments, the algorithm will first predict whether they are interacting or not. Using proteins predicted to interact, the algorithm will then predict contacts between columns in the alignment. These two steps are repeated until convergence is reached. This approach is still being tested.	Proteins poster	Fundamental
P_Pr070	845	Sudad Dayl and Ralf Schmid	Ralf Schmid	Structure prediction of the human P2X1 receptor using a homology modelling, ab initio modelling and cross-linking approach	P2X receptors are trimeric ion channels that are activated by the binding of ATP. Each P2X subunit consists of a large extracellular loop, two transmembrane helices, and intracellular amino and carboxy termini. In vertebrates, there are seven genes coding for P2X receptor subtypes. In particular, P2X1 and P2X7 receptors are drug targets for pain management, so structural information for human P2X1 and P2X7 receptors is of great interest. X-ray structures of the zebrafish P2X4 receptor in the closed state and the open state with ATP bound enhanced our understanding of this enigmatic family of ion channel receptors. However, the C and N terminal regions which range from ~ 24-30 and 27-240 residues, respectively were not present in the constructs used for crystallization. To gain insight into the structure of the human P2X1 receptor, we applied a hybrid modelling approach. The extracellular domain and TM helices were homology modelled based on the zP2X4 template (44 % sequence identity). This was combined with fragment-based ab initio prediction for the 20 N-terminal and 20 C-terminal residues of the intracellular domain using ROSETTA with symmetry constraints and anchoring in the membrane. After clustering 10 groups of alternative models were obtained. These clusters of models are validated by site-directed mutagenesis and crosslinking.	Proteins poster	Fundamental
P_Pr071	702	Michael Ringel and Thomas Brüser	Michael Ringel	SubtleP - A new software for subcellular translocation & localization prediction	Protein translocation systems are important for the interaction of microorganisms with their surroundings, especially in host-microbe interactions for instance during infections or in symbiotic-, parasitic- or commensalistic relations. Thus the prediction of these protein translocation systems and their respective substrates might shed light on their functional relevance and facilitate phenotypic screening. Furthermore, by unraveling the functional significance, new targets for antimicrobial drugs may be identified.The identification of protein translocation systems and their respective substrates poses a major challenge for bioinformatics and many algorithms have been devised over the last years to solve this problem. Recently meta-predictors have been developed, which address individual strengths and weaknesses of these algorithms, thus optimizing prediction-accuracy. Due to the multitude of available algorithms and their relevant technicalities it may often be difficult to combine the obtained results and to evaluate their significance in the abovementioned research topics. Therefore, we designed a new user-friendly software taking advantage of many open-source, published prediction algorithms to make these available for proteome wide predictions with an interactive graphical output. This may facilitate the usage of said algorithms by a broader audience. Moreover, the software has been designed in a modular fashion, making well established algorithms available to developers as building-blocks and abstracting basic tasks such as parsing files. Therefore, developers may assemble their own predictions-algorithms upon this infrastructure, thus expediting software development in this field.	Proteins poster	Fundamental
P_Pr072	686	Bálint Mészáros, András Zeke, Attila Reményi, István Simon and Zsuzsanna Dosztányi	Bálint Mészáros	Systematic analysis of somatic mutations driving cancer: Uncovering functional protein regions in disease development	Recent advances in sequencing technologies enable the large-scale identification of genes that are affected by various genetic alterations in cancer. However, understanding tumor development requires insights into how these changes cause altered protein function and impaired network regulation in general and/or in specific cancer types.In this work we present a novel method called iSiMPRe [1] that identifies regions that are significantly enriched in somatic mutations and short in-frame insertions or deletions (indels). Applying this unbiased method to the complete human proteome, by using data enriched through various cancer genome projects, we identified around 500 protein regions which could be linked to one or more of 27 distinct cancer types. These regions covered the majority of known cancer genes, surprisingly even tumor suppressors. Additionally, iSiMPRe also identified novel genes and regions that have not yet been associated with cancer.While local somatic mutations correspond to only a subset of genetic variations that can lead to cancer, our systematic analyses revealed that they represent an accompanying feature of most cancer driver genes regardless of the primary mechanism by which they are perturbed during tumorigeneses. These results indicate that the accumulation of local somatic mutations can be used to pinpoint genes responsible for cancer formation and can also help to understand the effect of cancer mutations at the level of functional modules in a broad range of cancer driver genes.[1] Mészáros B, Zeke A, Reményi A, Simon I, Dosztányi Z.Biol Direct. 2016 May 5;11:23. doi: 10.1186/s13062-016-0125-6. PMID: 27150584	Proteins poster	Health
P_Pr073	473	Dániel Györffy, Péter Závodszky and András Szilágyi	Dániel Györffy	The blind leading the blind: how disordered peptides form an ordered complex	Disordered proteins lack a well-defined three-dimensional structure in their free form in solution but can go through a disorder-to-order transition when binding to their cellular targets. When two disordered proteins form a complex—for example a homodimer—both molecules can become ordered. Because of the huge number of degrees of freedom of a system consisting of two disordered proteins, the computational description of such systems is a serious challenge. We have introduced a two-layer network model to describe the kinetics and the mechanisms of the coupled folding and binding processes occurring during the homodimer formation of disordered peptides. In contrast to the two mechanisms used for the description of ligand binding of proteins, namely induced fit and conformational selection, we distinguish three possible scenarios for the homodimer formation of disordered proteins: (i) rigid docking, where both molecules become ordered before association, (ii) induced fit, where association occurs before the folding of either monomer, and (iii) a mixed type where one of the chains is unfolded while the other is folded when association takes place. Applying our two-layer network model to 2D HP latticemodel dimers andWako–Saitо–Muñoz–Eaton models of several known protein dimers with different experimental behaviors, we found that dimer formation occurs via all three mechanisms for all sequences. The contribution of each mechanism depends on the particular sequence, the type of process (equilibrium or steady-state), and can even vary in time. These results also indicate that dimer formation can proceed by different mechanisms in vivo than in vitro.	Proteins poster	Fundamental
P_Pr074	635	Diego Honda, Sônia Freitas and João Martins	Diego Honda	The Bowman-Birk inhibitor from Vigna unguiculata seeds (BTCI) in complex with Trypsin: a molecular orbital study	BTCI is a Bowman-Birk Trypsin/Chymotrypsin inhibitor from Vigna unguiculata seeds with high biotechnological potential, especially due to its pharmacological characteristics. It presents seven disulfide bonds, which are responsible for its high stability in a broad range of temperature and pH conditions. In this context, it was chosen three semi-empirical methodologies to get chemical insights on structure of the BTCI-trypsin interface and its relationship with inhibition process. To accomplish this objective, we explore the frontier orbitals and their four immediate neighbors. In order to understand the local interactions, we also studied the BTCI and trypsin in vacuum. Likewise, the energy of each disulfide bond of the BTCI was determined. We obtained different behavior for each methodology for trypsin and BTCI, and the BTCI-trypsin complex. However, when we analyzed the interface between those two proteins, all methods are in agreement, pointing out that Cys22 is responsible to maintain the interface conformation during the enzyme-inhibitor interaction.	Proteins poster	Biotechnology Fundamental Health
P_Pr076	632	Flavia Corsi, Alessandra Carbone and Elodie Laine	Flavia Corsi	Towards an accurate prediction of protein-DNA interfaces based on evolutionary information, physico-chemical properties of residues and local geometry of the protein structure.	Protein interactions are essential to all biological processes and they represent increasingly important therapeutic targets. A new method was recently developed for accurately predicting protein-protein interfaces, understanding their properties, origins and binding to multiple partners (Laine & Carbone, PLoS Comp. Biol. 2015). This combines in a rational and very straightforward way three sequence- and structure-based descriptors of protein residues: evolutionary conservation, physico-chemical properties and local geometry. The implemented strategy yields very precise predictions for a wide range of protein-protein interfaces and discriminates them from small-molecule binding sites, permitting to dissect interaction surfaces. The approach is implemented in JET2, an automated tool for sequence-based protein interface prediction that is based on the Joint Evolutionary Trees (JET) method (Engelen … Carbone, PLoS Comp. Biol. 2009).We developed new strategies for predicting protein interfaces involved in protein-DNA interactions. These interaction surfaces are expected to satisfy characteristics different from those of protein-protein interfaces. We analyzed the evolutionary conservation, physico-chemical and geometrical properties of protein-DNA interfaces and we observed that not only physico-chemical properties but also geometrical patterns holding for protein-protein interactions are not anymore true for DNA-protein interactions. Then, by approaching the question as in JET2, we defined a few new rational heuristics leading to accurate protein-DNA interface identifications.This analysis (based on geometrical descriptors) can be used as the basis for the development of an optimal model of protein-DNA interaction. Other directions are constituted by RNA-protein interaction and small molecules-protein interactions, already partially addressed when analyzing protein-protein interactions (Laine & Carbone, 2015).	Proteins poster	Fundamental
P_Pr077	821	Julia Varga, Laszlo Dobson, Istvan Remenyi and Gabor E. Tusnady	Julia Varga	TSTMP: Target Selection for human TransMembrane Proteins	Transmembrane proteins (TMP) play an important role in living cells, since they are involved in diverse biological processes. Despite the great striving of worldwide structural genomics centres of membrane proteins, there are only around 60 known 3D structures among the human transmembrane proteins (with 2 or more transmembrane segments) and a further 600-700 could be modeled using existing structures. TSTMP database is a resource of human transmembrane proteins considering the existence of an exact 3D structure, or the possibility of modeling structure for the protein using existing 3D structure, or the necessity of a new structure for modeling the protein. The database was built by sorting out proteins from the human transmembrane proteome [1] with known structure and searching for suitable model structures for the remaining proteins by combining the results of state-of-the-art transmembrane specific fold recognition[2] and sequence similarity search[3] algorithms. TMPs were searched for homologues among the human transmembrane proteins to select targets whose successful structure determination would lead to the best structural coverage of the human transmembrane proteome. The database is available at http://tstmp.enzim.ttk.mta.hu.[1] The Human Transmembrane ProteomeDobson, L., Reményi, I. and Tusnády, GE: (2015)Biology Direct, 10:31[2] TMFoldRec: a statistical potential-based transmembrane protein fold recognition tool.Kozma, D. and Tusnády, GE. (2015)BMC Bioinformatics, 16, 201.[3] HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignmentRemmert, M., Biegert, A., Hauser, A. and Söding, J. (2011)Nat. Methods, 9, 173–175.	Proteins poster	Biotechnology

P_Pr078	463	Aram Gyulkhandanyan	Aram Gyulkhandanyan	Two paths of tumors destruction	Currently destruction of cancer cells actively studied in two directions: (i) by method of photodynamic therapy (PDT) and (ii) by acting on receptors of cancer cells leading to prevention of their dimerization. These studies are carried out both via experimental methods and the method of computer modelling (molecular docking). (i) As a damaging agent in a method of PDT are used photosensitizers (usually porphyrins). Photosensitizers accumulate selectively in tumors and upon illumination promoted generating of reactive oxygen species in cells that result to the destruction of cancer cells. (ii) The epidermal growth factor receptor (EGFR) is a membrane-spanning protein that governs major signaling pathways, as a result of its over expression and deregulation goes an aggressive tumor growth. Together with scientists from the University of Nantes we have shown that some small compounds (non-peptide compound nitro-benzoxadiazolyl (NBD)) may bind to dimerization domain EGFR. This causes allosteric activation of receptor, promotes the formation of stable dimers and launching of oncological process. On the other hand via molecular docking and spectral methods we showed high affinity of cationic porphyrins with a number of proteins (serum albumin, hemoglobin, cytochrome c), as well as with low molecular weight compounds (long chain fatty acids). It allows assuming that at complexation of porphyrins (with EGF or with the molecule type NBD) with the extracellular dimerization domains I and III of EGFR and by photodynamic illumination, the active oxygen species can cause destruction of the domains, prevent the dimerization process and cancer launching.	Proteins poster	Fundamental
P_Pr079	700	Erzsébet Fichó, Bálint Mészáros and István Simon	Erzsébet Fichó	Two-state Protein Complexes	Intrinsically disordered proteins (IDPs) lack a well-defined 3D structure. Their disordered nature enables them to fulfill several vital biological roles. Among others they participate in transcription, cell signaling, regulation, and stress-response. Disordered proteins rarely act alone; they are key elements of protein-protein interaction networks, often playing roles in signal transduction. In recent years it became clear that many IDPs are involved in disease development.Protein complexes formed by ordered proteins are well studied; however, the growing number of known disordered proteins and their functions require us to analyze interactions in ordered-disordered and disordered-disordered complexes. While ordered-disordered complexes have also been studied in details in recent years, the “two-state” (disordered-disordered) complexes remain a grey area of protein interactions.These type of interactions are unique because the complexes are ordered, while all participating proteins are unstable when separated. Although, these interactions are vital for the living cells, as of yet there is no available database to collect them. One of our aims is to identify these interactions through bioinformatical approaches. In order to organize our verified results, and to provide a starting point for further research on the background of these two-state protein systems, we also intend to build an online database. Deep analysis of the dataset can lead us to a better understanding of two-state protein complexes. As a long-term objective, it can provide novel pharmaceutical approaches, and can expand our knowledge of pathways mechanisms.	Proteins poster	Fundamental
P_Pr080	332	Alexandre Renaux, Ricardo Antunes, Cecilia Arighi, Andrea Auchincloss, Delphine Baratin, Alan Bridge, Elisabeth Coudert, Béatrice Cuche, Edouard De Castro, John S. Garavelli, Emma Hatton-Ellis, Guillaume Keller, Kati Laiho, Maria Martin, Alistair MacDougall,	Alexandre Renaux	UniRule - Increasing Annotation Depth of Unreviewed Protein Entries in UniProtKB.	UniProt provides a comprehensive and thoroughly annotated protein resource to the scientific community, most notably through the UniProt Knowledgebase (UniProtKB). Within UniProtKB, the reviewed section (Swiss-Prot) contains high quality, manually curated, richly-annotated protein records. In contrast, the unreviewed section (TrEMBL) which makes up 99% of UniProtKB, depends for its annotation on links to other databases and rule-based annotation systems. The use of rule-based annotation is necessary because there is no experimental data available for the majority of the unreviewed protein sequences.UniRule is a rule-based annotation system leveraging the expert-curated data in reviewed UniProtKB to increase the depth of annotation in unreviewed entries. Currently the UniRule system contains over 4,500 rules, which provide annotation for approximately 28% of unreviewed entries.Rules are a formalized way of expressing an association between conditions, which have to be met, and annotations, which are then propagated. InterPro signatures, predictive models for the functional classification of protein sequences, and taxonomic constraints are the fundamental conditions that are used. As a result, UniRule enriches the functional annotation of proteins with nomenclatures, catalytic activities, Gene Ontology terms and sequence features such as transmembrane domains. Data provenance is documented using Evidence Ontology tags.A key feature of the UniRule curation tool is a statistical quality control system which allows curators to evaluate their rules against the reviewed entries, to make sure rules are as accurate as possible. A dedicated space on the uniprot.org website has recently been created to allow users to view and explore UniRule.	Proteins poster	Fundamental
P_Pr082	446	Aytug Kiper, David Ramirez, Susanne Rinné, Wendy Gonzalez and Niels Decher	David Ramirez	Why Kv1.5 blockers preferentially inhibit TASK-1 channels?	Atrial fibrillation and obstructive sleep apnea are responsible for significant morbidity and mortality in the industrialized world. There is a high medical need for novel drugs against both diseases, and here, Kv1.5 channels have emerged as promising drug targets. In humans, TASK-1 has an atrium-specific expression and TASK-1 is also abundantly expressed in the hypoglossal motor nucleus. We asked whether known Kv1.5 channel blockers, effective against atrial fibrillation and/or obstructive sleep apnea, modulate TASK-1 channels. Therefore, we tested Kv1.5 blockers with different chemical structures for their TASK-1 affinity, utilizing TEVC-recordings in Xenopus oocytes. Despite the low structural conservation of Kv1.5 and TASK-1 channels, we found all Kv1.5 blockers analyzed to be even more effective on TASK-1 than on Kv1.5. For instance, the IC50 values of AVE0118 and AVE1231 (A293) were 10- and 43-fold lower on TASK-1. To describe this phenomenon on a molecular level, we used in silico models and identified unexpected structural similarities between the two drug binding sites. Kv1.5 blockers, like AVE0118 and AVE1231, which are promising drugs against atrial fibrillation or obstructive sleep apnea, are in fact potent TASK-1 blockers. Accordingly, block of TASK-1 channels by these compounds might contribute to the clinical effectiveness of these drugs. The higher affinity of these blockers for TASK-1 channels suggests that TASK-1 might be an unrecognized molecular target of Kv1.5 blockers effective in atrial fibrillation or obstructive sleep apnea 1.1.Kiper, A. K. et al. Pflugers Arch. 467, 1081–1090 (2015).	Proteins poster	Health