

22.4

Monitoring of Protein-Protein Interactions in P450-containing Systems in Real Time

Yuri Ivanov¹, Sergey Usanov², and Gaston Hui Bon Hoa³

¹Institute of Biomedical Chemistry RAMS, Pogodinskaya 10 str, Moscow 119992, Russia; ²Institute of Bioorganic Chemistry, NAS of Belarus, Minsk; and ³Institut de Biologie Physico Chimique, Paris, France

This study is concerned with the real-time optical biosensor (Affinity Sensors, United Kingdom) monitoring of protein-protein interactions in multi-protein systems: a water soluble cytochrome P450cam monooxygenase system containing only water-soluble proteins, cytochrome P450cam (P450cam), putidaredoxin reductase (PdR) and putidaredoxin (Pd); a membrane cytochrome P4502B4 system containing only membrane proteins, cytochrome P450 reductase (d-Fp), cytochrome P4502B4 (d-2B4) and cytochrome b5 (d-b5); and a mixed P450scc system containing cytochrome P450scc (P450scc), adrenodoxin reductase (AdR) and adrenodoxin (Ad). The functioning of these cytochrome P450 systems is realized by means of interprotein electron transfer from the donor protein, PdR or AdR or d-Fp to the acceptor protein, P450cam or P450scc or d-2B4 through the intermediate protein, Pd or Ad or d-b5 within the P450cam; or P450scc or d-2B4 systems, respectively; or directly - from d-Fp to d-2B4. The analysis of kinetics of complex formation and decay showed that electrostatic interactions play a dominant role in formation of electron-transferring binary complexes within water-soluble-protein-containing systems i.e. in formation of such complexes as Pd/PdR and Pd/P450cam, as well as Ad/AdR and Ad/P450scc. Unlike in the P450cam and P450scc systems, in the membrane P4502B4 system containing only membrane proteins, the hydrophobic interactions appear to be the driving force behind the formation of electron-transferring complexes, d-Fp/d-2B4 and d-b5/d-2B4. In all the three systems studied, there were also revealed the ternary complexes, PdR/Pd/P450cam, AdR/Ad/P450scc and d-Fp/d-2B4/d-b5, whose lifetimes were long enough to permit realization of complete substrate hydroxylation cycles.

This work was supported by RFBR Grants N 01-04-48245, 02-04-49057, 00-15-97926.

22.5

High-throughput Mapping of Protein-Protein Interactions: Automation of the Yeast Two-Hybrid System

Steve A. Richmond, Kelly Gregan, Kevin Russell, Chris Mann, and Julian F. Burke

Genetix Ltd., New Milton, Hampshire, BH25 5NN, United Kingdom

As we move into the post-genomic era, considerable academic and commercial interest is developing in elucidating the complex patterns of protein-protein interactions that are involved in both the function and regulation of virtually all biological processes. As the number of proteins encoded by the human genome is estimated to be up to one million, this is a formidable task that will only be achieved by the use of automated high-throughput techniques. The yeast two-hybrid system is an *in vivo* method for detecting protein-protein interactions that is particularly amenable to automation. As the interactions are detected in a eukaryotic context, it is especially suited to the study of mammalian proteins. At Genetix, we have developed an automated solution, the MegaMate, for carrying out high-throughput (over 100,000 clones per day) yeast two-hybrid screens. A 384-pin gridding tool is used to set up the matings on agar resulting in an addressable array of colonies. The gridding tool is then used to transfer the resulting colonies to selective media. The on-board imaging system uses a CCD camera to detect interactors, and maps their positions within the array, thus allowing full data tracking. All the data can be exported in XML format. The array format enables the incremental characterisation of previously uncharacterised libraries and opens up the possibility of systematically screening whole proteomes using each of the constituent proteins as a bait in order to map the entire complex of protein-protein interactions.

23.1

Database Resources for Proteome Research: SWISS-PROT, TrEMBL, InterPro

Rolf Apweiler

EMBL-EBI, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, United Kingdom

The progress in genomics and proteomics has led to large amounts of biological data that remains largely unexploited. *In silico* proteome analysis has become indispensable and complementary to genomic analysis. The EBI provides in collaboration with many partners key resources for the *in silico* analysis of completed proteomes by giving access to the complete sets of SWISS-PROT and TrEMBL protein sequences, InterPro based proteome-proteome comparisons and functional classification based on the GeneOntology. On these foundations the EBI will create in the EU funded Integr8 project in collaboration with partners all over Europe a gene-centric integration layer for complete genomes that allows to navigate from genome to transcriptome to proteome.

URLs:

<http://www.ebi.ac.uk/sprot/>
<http://www.ebi.ac.uk/interpro/>
<http://www.ebi.ac.uk/proteome/>
<http://www.ebi.ac.uk/GOA/>
<http://www.ebi.ac.uk/clustr/>

23.2

New Developments in Proteomics at the Swiss Institute of Bioinformatics

Ron D. Appel^{1,2,3}, Patricia Hernandez¹, Marc Tuloup¹, Daniel Walther^{1,2}, Robin Gras¹, Markus Müller¹, Christine Hoogland¹, Pierre-Alain Binz¹, Garry L. Corthals², Jean-Charles Sanchez², and Denis F. Hochstrasser^{2,3}

¹Swiss Institute of Bioinformatics, 1211 Geneva 4, Switzerland; ²Geneva Proteomics Centre, Geneva University Hospital, 1211 Geneva 14, Switzerland; and ³Geneva University, 1211 Geneva 4, Switzerland

Bioinformatics represents an essential part of proteomics. It encompasses a variety of aspects of the analysis of proteomes, including the use and development of large and comprehensive databases and knowledge bases, image analysis tools, laboratory information management systems, as well as software for protein identification and characterization from mass spectrometric data. The volume and complexity of proteomics data that are currently produced in many laboratories require the development of new, sophisticated tools. Some of them, currently under development at the Swiss Institute of Bioinformatics, will be presented.

23.3

Unity and Diversity of Protein Superfamilies

Semen Gusev and Alexander Archakov

Institute of Biomedical Chemistry, Pogodinskaya 10, 119992 Moscow, Russia

Bioinformatic analysis of protein sequences can elucidate the common and particular features in the groups of functionally related proteins. Here we report the results obtained for more than 2000 known enzymatic species of cytochromes P450 which are currently classified to 150 families and 250 subfamilies. Members of families and subfamilies share structural and functional similarity. Performed possesses information on the unique features of a certain families (diversity) and to establish the minimal structural requirements of a protein to be assigned to CYP superfamily (unity). Cluster analysis and multiple alignment of consensus sequences were used to produce the stable and optimal alignments for subfamilies, families and for the whole superfamily. Then motifs were revealed in the consensus sequences using the Sherman statistical analysis of distribution of conserved residues. Unity of cytochromes P450 is revealed at the superfamily consensus. Superfamily motifs (namely heme-peptide, meander, I- and K-helices) occur in every P450 protein and these motifs are likely to form the fold-determining core of P450s and to provide the function of monooxygenase catalysis. Diversity of P450 superfamily is defined by the motifs, which occur in family consensus and which are absolutely specific for a given family. These might be responsible for family catalytic peculiarities and for interaction with partner proteins. Revealed diversity motifs were mapped to the 3D structure of rabbit CYP2C5 and CYP51 of *M. tuberculosis*. Their participation in providing the substrate binding was demonstrated. Predictions were made concerning the families with no 3D models.

23.4

Statistical Standards for High Throughput Proteomics

Alexey I. Nesvizhskii, Andrew Keller, Jimmy K. Eng, Eugene Kolker, and Ruedi Aebersold

Institute for Systems Biology, Seattle, Washington

One of the priorities facing the proteomics community is the issue of data validation. This problem is especially acute in the case of high-throughput experiments performed on complex protein mixtures using LC/MS/MS strategy. Acquired MS/MS spectra are first searched against a sequence database using automated database searching programs such as SEQUEST or Mascot. Then, the database search results are used to derive a list of identified peptides and their corresponding proteins. This task necessarily entails distinguishing correct from false identifications among a large number of peptide assignments. We developed a robust and accurate statistical model to access the validity of peptide identifications made by MS/MS and database search. The method applies machine learning techniques to distinguish correctly from incorrectly assigned peptides in the data set and computes for each peptide identification a probability of being correct. These peptide probabilities are then used to estimate the likelihood that their corresponding proteins are present in the sample. The model is extensively tested using a test data set of MS/MS spectra generated from a sample of 18 control proteins. We demonstrate that using our approach one can achieve much higher sensitivity for any given error rate compared to the results of using conventional filtering criteria. The method enables high throughput analysis of proteomics data by eliminating the need to manually validate database search results. In addition, it can facilitate the benchmarking of various experimental procedures and serve as a common standard by which the results of different experimental groups should be compared.

23.5

Study of Spot Overlapping in 2D-PAGE Maps by Bidimensional Autocovariance Function

Nicola Marchetti¹, Maria Chiara Pietrogrande¹, Francesco Dondi¹, Attila Felinger², and Pier Giorgio Righetti³

¹Department of Chemistry, University of Ferrara, Italy; ²Department of Analytical Chemistry, University of Veszprem, Hungary; and ³Department of Agricultural and Industrial Biotechnologies, University of Verona, Italy

It is well known that, because of the very high number of components in any biological sample, the resolution power of 2D-PAGE is often not large enough to separate each single protein. Under standard conditions the overlapping between two or more single component spots cannot be avoided. The aim of this work is to set up a new chemometric approach able to interpret the complexity of protein separation (2D-maps), in terms of number of components, single component abundance distribution and average spot width. It is based on the 2-D autocovariance function (2D-ACVF) and its some important properties. First of all, it is able to single out the separation order hidden in the complex map: it is related to specific protein structure. A theoretical model has been developed to derive a relationship between the shape of the 2D-ACVF surface and the number of single components contained in the map as well as to the average standard deviation value of the spots. Different models able to describe how the 2D-ACVF surface changes by changing the spatial and spot intensity distributions are presented. They differ for the statistical hypothesis on the random nature of the these distributions. In order to test and validate the theoretical approach, a simulation procedure was developed and the models were applied on experimental data coming from real 2D-PAGE.