

Mini review

THE INTERACTOME: PREDICTING THE PROTEIN-PROTEIN INTERACTIONS IN CELLSDARIUSZ PLEWCZYŃSKI* and KRZYSZTOF GINALSKI
Interdisciplinary Centre for Mathematical and Computational Modelling,
University of Warsaw, Pawińskiego 5a, 02-106 Warsaw, Poland

Abstract: The term Interactome describes the set of all molecular interactions in cells, especially in the context of protein-protein interactions. These interactions are crucial for most cellular processes, so the full representation of the interaction repertoire is needed to understand the cell molecular machinery at the system biology level. In this short review, we compare various methods for predicting protein-protein interactions using sequence and structure information. The ultimate goal of those approaches is to present the complete methodology for the automatic selection of interaction partners using their amino acid sequences and/or three dimensional structures, if known. Apart from a description of each method, details of the software or web interface needed for high throughput prediction on the whole genome scale are also provided. The proposed validation of the theoretical methods using experimental data would be a better assessment of their accuracy.

Key words: Protein-protein interactions, Protein complexes, Docking, PDB Database, Interactome, Protein interaction networks, Physical protein interactions

*Author for correspondence; e-mail: D.Plewczynski@icm.edu.pl, tel.: (+48 22) 554-08-39, fax: (+48 22) 554-40-801

Abbreviations used: 3D - three dimensional; BIND - Biomolecular Interaction Network Database; BLAST - Basic Local Alignment Search Tool; CAPRI - Critical Assessment of PRediction of Interactions; DBID - database of interacting domains; DIP - Database of Interacting Proteins; DNA - deoxyribonucleic acid; GO - Gene Ontology; HPRD - Human Protein Reference Database; KEGG - Kyoto Encyclopedia of Genes and Genomes; KO - KEGG Orthology; MetaBASIC - Bilaterally Amplified Sequence Information Comparison; MINT - Molecular Interaction Database; OPHID - Online Predicted Human Interaction Database; PDB - Protein Data Bank; RNA - ribonucleic acid

INTRODUCTION

Interactions between proteins are crucial for most of the molecular processes in cells. Therefore, identifying the full interaction repertoire, the so-called protein interaction network, will yield a better understanding of the cellular machinery on the molecular level. A combination of experimental and theoretical approaches is needed to accomplish this goal. Large-scale experiments allow massive amounts of data to be gathered and interacting proteins to be characterized, even if the data is partial or inaccurate. It is essential to carefully select the most important pairs of interacting proteins to properly interpret such experimental results, i.e. observed protein interaction networks.

The term Interactome describes the set of all molecular interactions in a cell, especially in the context of protein-protein interactions. The whole interactome is often presented as a graph with nodes denoting interacting partners and edges representing interactions. In order to properly understand the molecular interactions, one has to start by analyzing interacting pairs of proteins in terms of their structure and sequence. Recent advances in bioinformatics tools and resources allow for a wider view of a protein and its biological context or physico-chemical properties. Thanks to the structural genomic initiative of the last decade, more and more protein sequences and their three-dimensional structures are known, giving rise to large publicly available databases. Furthermore, many experimental laboratories are performing more detailed experimental analyses of protein functions, yielding small but specialized databases. The most important goal of bioinformatics is integrating these experimental resources, developing computational methods based on these databases, and providing access to this converted information to a wider research audience.

The integration of various resources into a single, publicly available meta-database would allow for the easy extraction of valuable biological information from the massive amount of accumulated data. Moreover, the recent advances in theoretical methodology have been a crucial step into a systematic understanding of the molecular machinery. Most methods for the prediction of protein-protein interactions use sequence information to train various supervised machine learning algorithms. Three-dimensional structures acquired from the Protein Data Bank (PDB) [1] are also used to improve the accuracy of predictions. Further advances are possible when the genomic context, and predicted or known functional annotation of any interaction partners are taken into account. The number of crystallographically solved (known) structures is much smaller than the number of protein sequences. Therefore, most protein-protein interaction prediction methods neglect the three-dimensional details, focusing instead on sequence analysis. Major advances in the field of protein structure prediction allow the introduction of a complementary approach that will use predicted 3D models together with sequence features in order to correctly pair potentially interacting proteins. For instance, the MetaBASIC algorithm

developed by Ginalski *et al.* [2] compares the sequence profiles of two proteins enriched by their predicted secondary structures. Such alignments of query sequences with proteins of known structure allow for quick mean resolution 3D model creation and optimization.

The prediction of protein-protein interactions is a difficult problem when an analysis of both the protein sequences and known three-dimensional structures is needed. There are at least two reasons for that. Firstly, the number of possible interactions that have to be taken into account is extremely large. In yeast cells, one would have to analyze a possible 18,000,000 interactions between 6,000 proteins encoded in its genome, not counting multiple variants of gene products coming from alternative splicing or post-translational modifications. Only a small portion of those pairs is actually present in cells [3, 4]. Secondly, there are several types of possible interactions in living cells, from stable complexes up to temporary, functional pairings (for example, during the phosphorylation process in response to external stimuli). Protein complexes are better preserved during the evolution process than single proteins, so some computational methods focus on the prediction or searching of complexes that are common to several species [5-12]. Those methods use available information about experimentally verified interactions between proteins, orthologies, and comparison of protein sequences [13].

This review outlines recent advances in the field of protein-protein interaction prediction, and introduces some ideas about the application of structural modeling to improve the accuracy of these approaches. Such rapid methods allow the whole proteomes of various organisms to be scanned. By improving selection accuracy, computational methods help in the discovery of true interactions within the experimental data, which contains many false positives. Thus, the ultimate goal of system biology, i.e. to draw and understand the real interactome, is close to being achieved.

MATERIALS AND METHODS

Databases

The main goal for scientists of post-genomic biology is understanding the complex network of interacting proteins, DNA, RNA and small chemical molecules in living cells. Proteins are crucial parts of those networks, and information about their structure, amino acid sequence and functional context is important for gaining a better insight into the whole cell. Furthermore, knowledge about the interaction partners of a selected protein helps to obtain a more detailed description or even prediction of that protein's function [14, 15]. The gathered information about proteins for various species can be presented as complex graphs. Each edge can be described by a weight estimating the likelihood of a certain interaction. Two types of interactions can be distinguished: physical coupling (where stable or meta-stable complexes are formed) or functional coupling (where the pair of proteins is connected due to the mutual or directed

influence of one on another during a chemical reaction). Both types of interaction are very different in terms of the molecular details, but they can be integrated into the same complex graph. For example, the signaling cascades can be displayed as linear graphs of interacting proteins, where each node modifies the next one to transmit a signal. Such chains represent simple paths in the complex networks of interactions that also contain the stable protein complexes [16]. Several sources of biological information on protein sequences, structures and their interactions are available online. Those resources can be roughly divided into two groups: sequence and structural. The first group of experimentally confirmed protein-protein pairings involves transient interactions, and the second focuses on complexes, i.e. stable interactions. Most of the databases use their own format for the data, so their current level of integration is limited. The theoretical analysis of interactions depends on heterogeneous sources of biological information, such as sequence and structural databases, the literature, and experimental data. The main databases containing experimental information about protein-protein interactions are: the Database of Interacting Proteins (DIP) [17], the Biomolecular Interaction Network Database (BIND) [18], the Molecular Interaction Database (MINT) [19], INTACT [20, 21], and Human Protein Reference Database (HPRD) [22]. The literature data on selected protein sequences is available from the iHOP [23] and STRING [24] databases. In the Protein Data Bank (PDB) [1] database, one can find the three-dimensional structures of protein complexes, while in SCOP, one can find protein domains, whereas protein families in PFAM [25]. It is possible to find the homologous proteins to each of a given protein's interacting partners using tools like PSI-BLAST [26]. The PSI-PRED [27] allows the secondary structure of a given protein to be predicted using only sequence information, and can be used to enrich the sequence data by some structural features. In a similar way, one can assign the likely molecular function through functional categories from COG [28], GO annotations [29], the GOA database [30, 31], or the metabolic classification KO [32]. In order to enrich this information, it is useful to include information about the participation of a query protein in the metabolic or signaling pathways using the standard toolbox from the KEGG <http://www.genome.jp/kegg/> [33] and GO <http://www.geneontology.org/> [29] consortia.

The whole set of available sources of biological information on interacting protein sequences, and experimental or predicted three-dimensional complexes could be integrated into a single meta-database. Meta-databases allow for significant advances in the analysis of the statistical properties of protein interaction networks. The yeast analysis by Jeong *et al.* [34] or Sprinzak *et al.* [35] can be used as an example. This organism is well described in terms of experimental data, which allows for a detailed analysis of transient and stable interactions between the proteins of the yeast proteome. Another example of such a meta-database is OPHID (the Online Predicted Human Interaction Database) [36]. This resource only contains human proteins compiled from

BIND, HPRD and MINT, along with predictions based on interaction networks from yeast, fruit fly, mouse and other species, with the underlying assumption that orthologous proteins have similar interaction partners.

Such integrated data was used to propose a novel theoretical method of protein-protein interaction prediction using logistic regression on a selected set of global protein attributes. This example clearly shows that the availability of high fidelity data is of crucial importance for the further development of theoretical methods.

Algorithms

Detailed experimental data is needed for a better understanding of the functional rules governing cellular life on the molecular level. Over the last decade, we could observe significant progress in the experimental techniques for the identification of interactions between proteins. Several types of experimental assay, such as the yeast two-hybrid assay [37-40] or tandem affinity purification [41], allow the high efficiency experimental analysis of protein-protein interactions on the whole proteome scale (for a review of the experimental methods, please refer to [11, 12, 42-52]).

Those advances in experimental methodology quickly led to progress in theoretical approaches, such as those based on homology [53], protein pathways analysis [54], multimeric threading [55], or the prediction of interaction sites by docking methods [56]. The latter depends on previously acquired three-dimensional protein structures [57]. New data from high throughput methods of structural genomics allow for further advances in this field. The interacting sites on protein surfaces are often hydrophobic [58, 59], with evolutionarily conserved polar residues, so-called “hot spots” [60-62]. The important factor influencing the accuracy of interaction predictions is the proper representation of the proteins.

There are two classes of input data used in most theoretical methods: *sequence profiles* (phylogenetic profiles [63], mRNA expression levels [64], the presence of protein domains in analyzed sequences [65], and other approaches [66-68]); and *three-dimensional structures* of interacting partners (prediction methods tested every year in independent computational experiments in the Critical Assessment of PRediction of Interactions, CAPRI [57], methods based on the features of protein surfaces that allow for typing of the protein complexes [58, 69-71], the selection of the most important residues and their features on protein interfaces [59, 72-74], and other representations of experimental data [6, 75-77]). If the complex structure is known, it is possible to select interacting interfaces on the protein surfaces, find the most important residues, and analyze their evolutionary conservation or physico-chemical features [6, 58, 78, 79]. Those methods do not depend strongly on the details of the protein structures, because the evolutionary information of each amino acid is projected from multiple sequence alignments mapped onto a selected, single protein structure [6, 80, 81]. Unfortunately, such an approach has some inherent limits, as

described recently [82, 83]. The accuracy of these methods is limited to a few experimentally characterized protein families. We are aware of only a few experiments that have confirmed those predictions.

Lately, new methods emerged utilizing the three-dimensional structures of single protein crystals compared to their structures in the crystals of complexes [84]. Both spatial descriptors of the solved proteins are analyzed in terms of some structural features (such as solvent accessibility) that are selected by machine learning algorithms in order to build representations that well distinguish interacting and non-interacting partners.

The current theoretical and experimental approaches have different characteristics of systematic errors, and focus on different parts of the whole interactome. In most cases, one can observe only a small intersection between the various data sources, with no single method providing the perfect accuracy and good selectivity of predicted interaction networks. It is also important to stress that interactions confirmed by two or more methods are much more likely to be true than those identified by only a single method [3, 4]. This observation gives rise to the whole family of meta-predictors, i.e. methods that use several independent algorithms and experimental data sets in order to predict interactions. For data analysis, most of these use various machine learning methods, such as the support vector machine [66, 83, 85], naïve Bayesian classification [86], and decision trees [87, 88]. All those methods use only sequence information and single machine learning methods, and do not build consensus between various machine learning algorithms.

The methods use sequences of proteins or their known three-dimensional structures in order to predict interactions. In most cases, the three-dimensional structures of the interacting molecules are not known. Therefore, most of the existing methods focus on sequence-based protein-protein interaction predictions, and use interacting sequences as their training sets. Knowledge about complex structures helps in the selection of positives and negatives for training, which is of crucial importance for machine learning methods. Zhou and Shan [89] predicted interactions between proteins using neural networks trained on the sequence profiles of interacting proteins for 615 pairs of non-homologous proteins and predicted solvent accessibility residues. In independent tests of 129 pairs of proteins, their method predicts that 70% of 11,004 residues take part in the formation of the complex. The list of neighbor residues on a protein chain and solvent accessibility are not dependent on structural changes during complex formation. Therefore, the accuracy of the method is not worse when single protein crystals are used instead of three-dimensional complex structures. A similar approach was proposed by Fariselli *et al.* [6], who focused on the selection of important structural features from known protein complexes. Their neural network is able to predict 73% of the interaction residues on the independent benchmark of 226 structures. This work confirms that the use of physico-chemical features of interaction patches and sequence profiles allows for the proper selection of residues important for interactions.

Interacting residues tend to build clusters in protein sequences, so the support vector machine can easily be applied [83]. Ofran and Rost [68] proposed a simple neural network that can predict which protein fragments interact. The 94% high fidelity predictions for 34 of 333 proteins were confirmed experimentally. In the case of good predictions, almost 70% are confirmed experimentally, and at least one interaction patch is predicted correctly in 20% of cases, i.e. in 66 of 333 protein complexes. Those results show that effective prediction using only sequence information is possible. The inclusion of evolutionary information and structural descriptors improves those methods.

RESULTS AND DISCUSSION

Recent advances in System Biology, especially in the context of high-throughput DNA sequencing (genomics), gene expression (transcriptomics), metabolite and ion analysis (metabolomics/ionomics) and protein analysis (proteomics) carries with it the challenge of processing and interpreting the accumulating data sets [90-95]. Publicly accessible databases and bioinformatic tools are employed to mine this data in order to filter relevant correlations and create models describing physiological states [90]. Reconstructing the networks of interactions of the various cellular components as enzyme activities and complexes, gene expression, metabolite pools or pathway flux modes is therefore possible. However, capturing the interactions of network elements requires experimental setups with a variety of conditions [90, 95-99]. The ultimate goal of systems biology in the context of plant research is to understand the molecular principles governing plant responses and consistently explain plant physiology [90, 100, 101]. These approaches were described in detail in this very journal in the 2001-2008 period.

Recent studies of protein-protein interactions focused on using some global sequence or structural features. For example Sprinzak *et al.* [35] presented a linear regression method trained on nine global protein attributes, such as domain signature, fold type, gene fusion, phylogenetic profile, gene context, conservation of neighboring genes, protein localization, type of molecular pathway, mRNA coexpression, or transcription coregulation. A protein could then be represented as points in nine-dimensional space using those features, and linear regression was applied. The problem is very complicated and non-trivial. Proper selection of the protein domain is necessary [102-108]. In addition to pure chemical data [109-116] in the context of the Drug Discovery [117-127], there is also a need for some knowledge on protein-protein interactions, the high quality structural prediction of proteins [2, 128-136] and their inhibitors, and a detailed understanding of how those inhibitors affect the molecular recognition between proteins.

The development of theoretical methods for function annotation clearly shows that a detailed analysis of local characteristics of the protein chain can significantly improve accuracy. Recently published papers [137-142] use local

sequence description with some structural features (for example relative solvent accessibility RSA) for much more efficient training of machine learning methods. In the work of Meller *et al.* [84], the authors tested the whole set of machine learning methods (neural networks, support vector machines and linear discrimination). Those methods were trained on the stable protein complexes from the Protein Data Bank (PDB). This work strongly supports the use of machine learning methods and the local representation of proteins. The authors obtained almost 70% accuracy of protein-protein interaction prediction, provided the structures of the interaction partners were known.

The next step in the development of computational methods is based on several structural features of both interacting partners predicted from a sequence. The sequences can be represented as sets of short fragments with a known homology profile (build for the whole proteins). Such an approach was first used by Ofra and Rost in 2003 [68], and later by Fernandez-Ballester and Serrano [143] as position-specific multiple-sequence alignment matrixes. Both broad types of interactions (stable and transient) are described as interactions between some key amino acids of both partners. The proper selection of important interacting residues allows for the performance of more certain predictions. Therefore, the analysis of contact maps in protein complexes with predicted local structural conformation of the main chain enriched by homology profiles will allow the collection of a more natural training set for machine learning methods. The proper representation of both the sequence and structure of the interacting partners is of crucial importance for the further development of bioinformatics computational methods for protein-protein interaction prediction [144, 145].

The sequence-based methods typically search for homologues of both interacting proteins with tools such as PSI-BLAST [26] or RPS-BLAST. If some interacting partners are present in both sets, the protein pair is likely to interact. Those methods are not yet sensitive enough in the case of distant homology. Some proteins from large and diverse superfamilies are dissimilar in terms of sequence: only a fold and a few important residues in an active site are preserved [146]. The lack of clear sequence similarity makes new family identification or function prediction more difficult. Protein-protein interaction prediction is also more complex when no close interacting homologs are known. Structure-based methods have a very limited area of usefulness due to their dependence on known structures of protein-protein complexes. Therefore, more sensitive sequence (eg.: Meta-BASIC at <http://basic.bioinfo.pl/> [2]) or structure prediction methods (like those coupled in the Protein Structure Prediction Meta Server <http://BioInfo.PL/Meta/> [133, 147, 148]) will be a significant improvement over standard sequence-based methods, even without knowledge of the exact three-dimensional structures of the interacting partners.

Other types of approach focus on predicting protein-protein interactions using protein-protein interaction network topology [149, 150]. In their approach, a spectral method derived from graph theory is used to select the topological substructures of protein-protein interaction networks that are biologically

relevant functional groups. The function of uncharacterized proteins is then assigned based on the classification of known proteins within topological structures [149]. The clusters can also be identified by an eigenmode analysis of the connectivity matrix of the protein-protein interaction network. Such functional clustering allows for further prediction of new protein interactions [150]. The structural matching is also used to recognize protein-protein interaction sites in protein structures [151]. Prism (<http://gordion.hpc.eng.ku.edu.tr/prism>) uses protein interface structures derived from the Protein Data Bank (PDB) for the automated prediction of protein-protein interactions. Both structure and sequence conservation in protein interfaces are employed providing additional insight into the protein-protein interaction predictions [152]. The InterPreTS server <http://www.russell.embl.de/cgi-bin/interprets2>, given a pair of query sequences, searches for homologs in a database of interacting domains (DBID) of known three-dimensional complex structures. Pairs of sequences homologous to a known interacting pair are then scored for how well they preserve the atomic contacts at the interaction interface [153-155].

Up to now, all the computational algorithms have only used single machine learning methods for the analysis and prediction of protein-protein interactions [156-160], or the statistical analysis of interacting patches of protein surfaces [75, 149, 161, 162]. Our experience clearly supports the idea that each machine learning algorithm performs better for selected types of training data [163, 164]. Some have very high specificity, others focus more on sensitivity. Sometimes one can have a very large number of positives in training, but it is also common for some specific types of experiments with only a few confirmed instances known. In most cases, the proper selection of negatives is not trivial. In the case of protein-protein interactions, one should use a rich variety of input data for training, such as sequences, short sequence motifs, evolutionary information, genomic context, enzymatic classification, or the known or predicted local or global structure of interacting proteins. Using this data, one can apply various types of machine learning methods trained on the same set of positives and negatives, for example neural networks, the support vector machine, the random forest, decision trees, or rough sets. The crucial step of meta-prediction is building a consensus between those various prediction methods. Since systematic errors of multiple methods are usually randomly distributed, the consensus approach can be used to select a common prediction, probably the most accurate one [165]. Thanks to its easy parallelization, a consensus method can improve the accuracy of any single machine learning method without extending the time of prediction (the time needed is equal to the slowest used machine learning algorithm). The combination of various approaches done by Sen and Kloczkowski [166] provides the solid justification for this statement. They combine four different methods, such as data mining using Support Vector Machines, threading through protein structures, prediction of conserved residues on the protein surface by analysis of phylogenetic trees, and the Conservatism of Conservatism method of Mirny and Shakhnovich [167-169]. A consensus

method predicts protein-protein interface residues by combining sequence and structure-based methods. Therefore, we hypothesize that consensus approaches are the main tools to handle the prediction of protein-protein interactions on the whole proteome level, the ultimate goal of system biology.

Acknowledgements. This study was supported by the EC BioSapiens (LHSGCT-2003-503265) 6FP project as well as the Polish Ministry of Education and Science (PBZ-MNIl-2/1/2005, N N301 159735 and N N301 159435), NIH (1R01GM081680-01), EMBO Installation and FNP (FOCUS) grants.

REFERENCES

1. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. The Protein Data Bank. **Nucleic Acids Res.** 28 (2000) 235-242.
2. Ginalski, K., von Grotthuss, M., Grishin, N.V. and Rychlewski, L. Detecting distant homology with Meta-BASIC. **Nucleic Acids Res.** 32 (2004) W576-W581.
3. Sprinzak, E., Sattath, S. and Margalit, H. How reliable are experimental protein-protein interaction data? **J. Mol. Biol.** 327 (2003) 919-923.
4. von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S.G., Fields, S. and Bork, P. Comparative assessment of large-scale data sets of protein-protein interactions. **Nature** 417 (2002) 399-403.
5. Carter, P., Lesk, V.I., Islam, S.A. and Sternberg, M.J. Protein-protein docking using 3D-Dock in rounds 3, 4, and 5 of CAPRI. **Proteins** 60 (2005) 281-288.
6. Fariselli, P., Pazos, F., Valencia, A. and Casadio, R. Prediction of protein-protein interaction sites in heterocomplexes with neural networks. **Eur. J. Biochem.** 269 (2002) 1356-1361.
7. Hoskins, J., Lovell, S. and Blundell, T.L. An algorithm for predicting protein-protein interaction sites: Abnormally exposed amino acid residues and secondary structure elements. **Protein Sci.** 15 (2006) 1017-1029.
8. Jothi, R., Cherukuri, P.F., Tasneem, A. and Przytycka, T.M. Co-evolutionary analysis of domains in interacting proteins reveals insights into domain-domain interactions mediating protein-protein interactions. **J. Mol. Biol.** 362 (2006) 861-875.
9. Tan, K., Shlomi, T., Feizi, H., Ideker, T. and Sharan, R. Transcriptional regulation of protein complexes within and across species. **Proc. Natl. Acad. Sci. USA** 104 (2007) 1283-1288.
10. Teichmann, S.A. Principles of protein-protein interactions. **Bioinformatics** 18 Suppl 2 (2002) S249.
11. Cusick, M.E., Klitgord, N., Vidal, M. and Hill, D.E. Interactome: gateway into systems biology. **Hum. Mol. Genet.** 14 Spec No. 2 (2005) R171-R181.

12. Goh, C.S. and Cohen, F.E. Co-evolutionary analysis reveals insights into protein-protein interactions. **J. Mol. Biol.** 324 (2002) 177-192.
13. Sharan, R., Ideker, T., Kelley, B., Shamir, R. and Karp, R.M. Identification of protein complexes by comparative analysis of yeast and bacterial protein interaction data. **J. Comput. Biol.** 12 (2005) 835-846.
14. Barash, Y., Elidan, G., Kaplan, T. and Friedman, N. CIS: compound importance sampling method for protein-DNA binding site p-value estimation. **Bioinformatics** 21 (2005) 596-600.
15. Sharan, R., Suthram, S., Kelley, R.M., Kuhn, T., McCuine, S., Uetz, P., Sittler, T., Karp, R.M. and Ideker, T. Conserved patterns of protein interaction in multiple species. **Proc. Natl. Acad. Sci. USA** 102 (2005) 1974-1979.
16. Kelley, B.P., Sharan, R., Karp, R.M., Sittler, T., Root, D.E., Stockwell, B.R. and Ideker, T. Conserved pathways within bacteria and yeast as revealed by global protein network alignment. **Proc. Natl. Acad. Sci. USA** 100 (2003) 11394-11399.
17. Salwinski, L., Miller, C.S., Smith, A.J., Pettit, F.K., Bowie, J.U. and Eisenberg, D. The Database of Interacting Proteins: 2004 update. **Nucleic Acids Res.** 32 (2004) D449-D451.
18. Alfarano, C., Andrade, C.E., Anthony, K., Bahroos, N., Bajec, M., Bantoft, K., Betel, D., Bobechko, B., Boutilier, K., Burgess, E., Buzadzija, K., Cavero, R., D'Abreo, C., Donaldson, I., Dorairajoo, D., Dumontier, M.J., Dumontier, M.R., Earles, V., Farrall, R., Feldman, H., Garderman, E., Gong, Y., Gonzaga, R., Grytsan, V., Gryz, E., Gu, V., Haldorsen, E., Halupa, A., Haw, R., Hrvojic, A., Hurrell, L., Isserlin, R., Jack, F., Juma, F., Khan, A., Kon, T., Konopinsky, S., Le, V., Lee, E., Ling, S., Magidin, M., Moniakakis, J., Montojo, J., Moore, S., Muskat, B., Ng, I., Paraiso, J.P., Parker, B., Pintilie, G., Pirone, R., Salama, J.J., Sgro, S., Shan, T., Shu, Y., Siew, J., Skinner, D., Snyder, K., Stasiuk, R., Strumpf, D., Tuekam, B., Tao, S., Wang, Z., White, M., Willis, R., Wolting, C., Wong, S., Wrong, A., Xin, C., Yao, R., Yates, B., Zhang, S., Zheng, K., Pawson, T., Ouellette, B.F. and Hogue, C.W. The Biomolecular Interaction Network Database and related tools 2005 update. **Nucleic Acids Res.** 33 (2005) D418-D424.
19. Chatr-Aryamontri, A., Ceol, A., Palazzi, L.M., Nardelli, G., Schneider, M.V., Castagnoli, L. and Cesareni, G. MINT: the Molecular INTeraction database. **Nucleic Acids Res.** 35 (2007) D572-574.
20. Hermjakob, H., Montecchi-Palazzi, L., Lewington, C., Mudali, S., Kerrien, S., Orchard, S., Vingron, M., Roechert, B., Roepstorff, P., Valencia, A., Margalit, H., Armstrong, J., Bairoch, A., Cesareni, G., Sherman, D. and Apweiler, R. IntAct: an open source molecular interaction database. **Nucleic Acids Res.** 32 (2004) D452-D455.
21. Kerrien, S., Alam-Faruque, Y., Aranda, B., Bancarz, I., Bridge, A., Derow, C., Dimmer, E., Feuermann, M., Friedrichsen, A., Huntley, R., Kohler, C., Khadake, J., Leroy, C., Liban, A., Lieftink, C., Montecchi-Palazzi, L.,

- Orchard, S., Risse, J., Robbe, K., Roechert, B., Thorneycroft, D., Zhang, Y., Apweiler, R. and Hermjakob, H. IntAct - open source resource for molecular interaction data. **Nucleic Acids Res.** 35 (2007) D561-565.
22. Peri, S., Navarro, J.D., Amanchy, R., Kristiansen, T.Z., Jonnalagadda, C.K., Surendranath, V., Niranjan, V., Muthusamy, B., Gandhi, T.K., Gronborg, M., Ibarrola, N., Deshpande, N., Shanker, K., Shivashankar, H.N., Rashmi, B.P., Ramya, M.A., Zhao, Z., Chandrika, K.N., Padma, N., Harsha, H.C., Yatish, A.J., Kavitha, M.P., Menezes, M., Choudhury, D.R., Suresh, S., Ghosh, N., Saravana, R., Chandran, S., Krishna, S., Joy, M., Anand, S.K., Madavan, V., Joseph, A., Wong, G.W., Schiemann, W.P., Constantinescu, S.N., Huang, L., Khosravi-Far, R., Steen, H., Tewari, M., Ghaffari, S., Blobe, G.C., Dang, C.V., Garcia, J.G., Pevsner, J., Jensen, O.N., Roepstorff, P., Deshpande, K.S., Chinnaiyan, A.M., Hamosh, A., Chakravarti, A. and Pandey, A. Development of human protein reference database as an initial platform for approaching systems biology in humans. **Genome Res.** 13 (2003) 2363-2371.
 23. Hoffmann, R. and Valencia, A. Implementing the iHOP concept for navigation of biomedical literature. **Bioinformatics** 21 Suppl 2 (2005) ii252-ii258.
 24. von Mering, C., Jensen, L.J., Snel, B., Hooper, S.D., Krupp, M., Foglierini, M., Jouffre, N., Huynen, M.A. and Bork, P. STRING: known and predicted protein-protein associations, integrated and transferred across organisms. **Nucleic Acids Res.** 33 (2005) D433-D437.
 25. Finn, R.D., Mistry, J., Schuster-Bockler, B., Griffiths-Jones, S., Hollich, V., Lassmann, T., Moxon, S., Marshall, M., Khanna, A., Durbin, R., Eddy, S.R., Sonnhammer, E.L. and Bateman, A. Pfam: clans, web tools and services. **Nucleic Acids Res.** 34 (2006) D247-D251.
 26. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. **Nucleic Acids Res.** 25 (1997) 3389-3402.
 27. Jones, D.T. Protein secondary structure prediction based on position-specific scoring matrices. **J. Mol. Biol.** 292 (1999) 195-202.
 28. Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V., Krylov, D.M., Mazumder, R., Mekhedov, S.L., Nikolskaya, A.N., Rao, B.S., Smirnov, S., Sverdlov, A.V., Vasudevan, S., Wolf, Y.I., Yin, J.J. and Natale, D.A. The COG database: an updated version includes eukaryotes. **BMC Bioinformatics** 4 (2003) 41.
 29. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M. and Sherlock, G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. **Nat. Genet.** 25 (2000) 25-29.

30. Camon, E., Barrell, D., Lee, V., Dimmer, E. and Apweiler, R. The Gene Ontology Annotation (GOA) Database - an integrated resource of GO annotations to the UniProt Knowledgebase. **In Silico Biol** 4 (2004) 5-6.
31. Camon, E., Magrane, M., Barrell, D., Binns, D., Fleischmann, W., Kersey, P., Mulder, N., Oinn, T., Maslen, J., Cox, A. and Apweiler, R. The Gene Ontology Annotation (GOA) project: implementation of GO in SWISS-PROT, TrEMBL, and InterPro. **Genome Res.** 13 (2003) 662-672.
32. Mao, X., Cai, T., Olyarchuk, J.G. and Wei, L. Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. **Bioinformatics** 21 (2005) 3787-3793.
33. Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y. and Hattori, M. The KEGG resource for deciphering the genome. **Nucleic Acids Res.** 32 (2004) D277-D280.
34. Jeong, H., Mason, S.P., Barabasi, A.L. and Oltvai, Z.N. Lethality and centrality in protein networks. **Nature** 411 (2001) 41-42.
35. Sprinzak, E., Altuvia, Y. and Margalit, H. Characterization and prediction of protein-protein interactions within and between complexes. **Proc. Natl. Acad. Sci. USA** 103 (2006) 14718-14723.
36. Brown, K.R. and Jurisica, I. Online predicted human interaction database. **Bioinformatics** 21 (2005) 2076-2082.
37. Cagney, G., Uetz, P. and Fields, S. High-throughput screening for protein-protein interactions using two-hybrid assay. **Methods Enzymol.** 328 (2000) 3-14.
38. Uetz, P., Giot, L., Cagney, G., Mansfield, T.A., Judson, R.S., Knight, J.R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., Qureshi-Emili, A., Li, Y., Godwin, B., Conover, D., Kalbfleisch, T., Vijayadamodar, G., Yang, M., Johnston, M., Fields, S. and Rothberg, J.M. A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. **Nature** 403 (2000) 623-627.
39. Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M. and Sakaki, Y. A comprehensive two-hybrid analysis to explore the yeast protein interactome. **Proc. Natl. Acad. Sci. USA** 98 (2001) 4569-4574.
40. Ito, T., Chiba, T. and Yoshida, M. Exploring the protein interactome using comprehensive two-hybrid projects. **Trends Biotechnol.** 19 (2001) S23-S27.
41. Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M. and Seraphin, B. A generic protein purification method for protein complex characterization and proteome exploration. **Nat. Biotechnol.** 17 (1999) 1030-1032.
42. Bader, G.D. and Hogue, C.W. Analyzing yeast protein-protein interaction data obtained from different sources. **Nat. Biotechnol.** 20 (2002) 991-997
43. Chen, T., Jaffe, J.D. and Church, G.M. Algorithms for identifying protein cross-links via tandem mass spectrometry. **J. Comput. Biol.** 8 (2001) 571-583.

44. Ito, T., Ota, K., Kubota, H., Yamaguchi, Y., Chiba, T., Sakuraba, K. and Yoshida, M. Roles for the two-hybrid system in exploration of the yeast protein interactome. **Mol. Cell. Proteomics** 1 (2002) 561-566.
45. McDermott, J., Bumgarner, R. and Samudrala, R. Functional annotation from predicted protein interaction networks. **Bioinformatics** 21 (2005) 3217-3226.
46. Morrison, J.L., Breitling, R., Higham, D.J. and Gilbert, D.R. A lock-and-key model for protein-protein interactions. **Bioinformatics** 22 (2006) 2012-2019.
47. Schweitzer, B., Predki, P. and Snyder, M. Microarrays to characterize protein interactions on a whole-proteome scale. **Proteomics** 3 (2003) 2190-2199.
48. Tong, A.H., Drees, B., Nardelli, G., Bader, G.D., Brannetti, B., Castagnoli, L., Evangelista, M., Ferracuti, S., Nelson, B., Paoluzi, S., Quondam, M., Zucconi, A., Hogue, C.W., Fields, S., Boone, C. and Cesareni, G. A combined experimental and computational strategy to define protein interaction networks for peptide recognition modules. **Science** 295 (2002) 321-324.
49. Walhout, A.J., Boulton, S.J. and Vidal, M. Yeast two-hybrid systems and protein interaction mapping projects for yeast and worm. **Yeast** 17 (2000) 88-94.
50. Wehr, M.C., Laage, R., Bolz, U., Fischer, T.M., Grunewald, S., Scheek, S., Bach, A., Nave, K.A. and Rossner, M.J. Monitoring regulated protein-protein interactions using split TEV. **Nat. Methods** 3 (2006) 985-993.
51. Wu, X., Zhu, L., Guo, J., Zhang, D.Y. and Lin, K. Prediction of yeast protein-protein interaction network: insights from the Gene Ontology and annotations. **Nucleic Acids Res.** 34 (2006) 2137-2150.
52. Yarmush, M.L. and Jayaraman, A. Advances in proteomic technologies. **Annu. Rev. Biomed. Eng.** 4 (2002) 349-373.
53. Marcotte, E.M., Pellegrini, M., Ng, H.L., Rice, D.W., Yeates, T.O. and Eisenberg, D. Detecting protein function and protein-protein interactions from genome sequences. **Science** 285 (1999) 751-753.
54. Troyanskaya, O.G., Dolinski, K., Owen, A.B., Altman, R.B. and Botstein, D. A Bayesian framework for combining heterogeneous data sources for gene function prediction (in *Saccharomyces cerevisiae*). **Proc. Natl. Acad. Sci. USA** 100 (2003) 8348-8353.
55. Lu, L., Lu, H. and Skolnick, J. MULTIPROSPECTOR: an algorithm for the prediction of protein-protein interactions by multimeric threading. **Proteins** 49 (2002) 350-364.
56. Smith, G.R. and Sternberg, M.J. Prediction of protein-protein interactions by docking methods. **Curr. Opin. Struct. Biol.** 12 (2002) 28-35.
57. Wodak, S.J. and Mendez, R. Prediction of protein-protein interactions: the CAPRI experiment, its evaluation and implications. **Curr. Opin. Struct. Biol.** 14 (2004) 242-249.
58. Jones, S. and Thornton, J.M. Analysis of protein-protein interaction sites using surface patches. **J. Mol. Biol.** 272 (1997) 121-132.

59. Lo Conte, L., Chothia, C. and Janin, J. The atomic structure of protein-protein recognition sites. **J. Mol. Biol.** 285 (1999) 2177-2198.
60. Glaser, F., Steinberg, D.M., Vakser, I.A. and Ben-Tal, N. Residue frequencies and pairing preferences at protein-protein interfaces. **Proteins** 43 (2001) 89-102.
61. Hu, Z., Ma, B., Wolfson, H. and Nussinov, R. Conservation of polar residues as hot spots at protein interfaces. **Proteins** 39 (2000) 331-342.
62. DeLano, W.L. Unraveling hot spots in binding interfaces: progress and challenges. **Curr. Opin. Struct. Biol.** 12 (2002) 14-20.
63. Pellegrini, M., Marcotte, E.M. and Yeates, T.O. A fast algorithm for genome-wide analysis of proteins with repeated sequences. **Proteins** 35 (1999) 440-446.
64. Eisen, M.B., Spellman, P.T., Brown, P.O. and Botstein, D. Cluster analysis and display of genome-wide expression patterns. **Proc. Natl. Acad. Sci. USA** 95 (1998) 14863-14868.
65. Sprinzak, E. and Margalit, H. Correlated sequence-signatures as markers of protein-protein interaction. **J. Mol. Biol.** 311 (2001) 681-692.
66. Bock, J.R. and Gough, D.A. Predicting protein-protein interactions from primary structure. **Bioinformatics** 17 (2001) 455-460.
67. Gallet, X., Charlotiaux, B., Thomas, A. and Brasseur, R. A fast method to predict protein interaction sites from sequences. **J. Mol. Biol.** 302 (2000) 917-926.
68. Ofran, Y. and Rost, B. Predicted protein-protein interaction sites from local sequence information. **FEBS Lett.** 544 (2003) 236-239.
69. Jones, S. and Thornton, J.M. Principles of protein-protein interactions. **Proc. Natl. Acad. Sci. USA** 93 (1996) 13-20.
70. Nooren, I.M. and Thornton, J.M. Diversity of protein-protein interactions. **Embo J.** 22 (2003) 3486-3492.
71. Nooren, I.M. and Thornton, J.M. Structural characterisation and functional significance of transient protein-protein interactions. **J. Mol. Biol.** 325 (2003) 991-1018.
72. Bahadur, R.P., Chakrabarti, P., Rodier, F. and Janin, J. A dissection of specific and non-specific protein-protein interfaces. **J. Mol. Biol.** 336 (2004) 943-955.
73. Ofran, Y. and Rost, B. Analysing six types of protein-protein interfaces. **J. Mol. Biol.** 325 (2003) 377-387.
74. Saha, R.P., Bahadur, R.P. and Chakrabarti, P. Interresidue contacts in proteins and protein-protein interfaces and their use in characterizing the homodimeric interface. **J. Proteome Res.** 4 (2005) 1600-1609.
75. Bordner, A.J. and Abagyan, R. Statistical analysis and prediction of protein-protein interfaces. **Proteins** 60 (2005) 353-366.
76. Neuvirth, H., Raz, R. and Schreiber, G. ProMate: a structure based prediction program to identify the location of protein-protein binding sites. **J. Mol. Biol.** 338 (2004) 181-199.

77. Chung, J.L., Wang, W. and Bourne, P.E. Exploiting sequence and structure homologs to identify protein-protein binding sites. **Proteins** 62 (2006) 630-640.
78. Valdar, W.S. and Thornton, J.M. Protein-protein interfaces: analysis of amino acid conservation in homodimers. **Proteins** 42 (2001) 108-124.
79. Yao, H., Kristensen, D.M., Mihalek, I., Sowa, M.E., Shaw, C., Kimmel, M., Kavraki, L. and Lichtarge, O. An accurate, sensitive, and scalable method to identify functional sites in protein structures. **J. Mol. Biol.** 326 (2003) 255-261.
80. Aloy, P., Querol, E., Aviles, F.X. and Sternberg, M.J. Automated structure-based prediction of functional sites in proteins: applications to assessing the validity of inheriting protein function from homology in genome annotation and to protein docking. **J. Mol. Biol.** 311 (2001) 395-408.
81. Berezin, C., Glaser, F., Rosenberg, J., Paz, I., Pupko, T., Fariselli, P., Casadio, R. and Ben-Tal, N. ConSeq: the identification of functionally and structurally important residues in protein sequences. **Bioinformatics** 20 (2004) 1322-1324.
82. Caffrey, D.R., Somaroo, S., Hughes, J.D., Mintseris, J. and Huang, E.S. Are protein-protein interfaces more conserved in sequence than the rest of the protein surface? **Protein Sci.** 13 (2004) 190-202.
83. Yan, C., Dobbs, D. and Honavar, V. A two-stage classifier for identification of protein-protein interface residues. **Bioinformatics** 20 Suppl 1 (2004) I371-I378.
84. Porollo, A. and Meller, J. Prediction-based fingerprints of protein-protein interactions. **Proteins** 66 (2006) 630-645.
85. Koike, A. and Takagi, T. Prediction of protein-protein interaction sites using support vector machines. **Protein Eng. Des. Sel.** 17 (2004) 165-173.
86. Jansen, R., Yu, H., Greenbaum, D., Kluger, Y., Krogan, N.J., Chung, S., Emili, A., Snyder, M., Greenblatt, J.F. and Gerstein, M. A Bayesian networks approach for predicting protein-protein interactions from genomic data. **Science** 302 (2003) 449-453.
87. Liu, X., Zhang, L.M. and Zheng, W.M. Prediction of protein secondary structure based on residue pairs. **J. Bioinform. Comput. Biol.** 2 (2004) 343-352.
88. Zhang, L.V., Wong, S.L., King, O.D. and Roth, F.P. Predicting co-complexed protein pairs using genomic and proteomic data integration. **BMC Bioinformatics** 5 (2004) 38.
89. Zhou, H.X. and Shan, Y. Prediction of protein interaction sites from sequence profile and residue neighbor list. **Proteins** 44 (2001) 336-343
90. Hesse, H. and Hoefgen, R. On the way to understand biological complexity in plants: S-nutrition as a case study for systems biology. **Cell. Mol. Biol. Lett.** 11 (2006) 37-56.
91. Hsieh, C.J., Chen, M.J., Liao, Y.L. and Liao, T.N. Polymorphisms of the uridine-diphosphoglucuronosyltransferase 1A1 gene and coronary artery disease. **Cell. Mol. Biol. Lett.** 13 (2008) 1-10.

92. Huang, B., Chu, C.H., Chen, S.L., Juan, H.F. and Chen, Y.M. A proteomics study of the mung bean epicotyl regulated by brassinosteroids under conditions of chilling stress. **Cell. Mol. Biol. Lett.** 11 (2006) 264-278.
93. Knizewski, L., Steczkiewicz, K., Kuchta, K., Wyrwicz, L., Plewczynski, D., Kolinski, A., Rychlewski, L. and Ginalski, K. Uncharacterized DUF1574 leptospira proteins are SGNH hydrolases. **Cell Cycle** 7 (2008) 542-544.
94. Korohoda, W. and Wilk, A. Cell electrophoresis - a method for cell separation and research into cell surface properties. **Cell. Mol. Biol. Lett.** 13 (2008) 312-326.
95. Li, J., Ji, C., Zheng, H., Fei, X., Zheng, M., Dai, J., Gu, S., Xie, Y. and Mao, Y. Molecular cloning and characterization of a novel human gene containing 4 ankyrin repeat domains. **Cell. Mol. Biol. Lett.** 10 (2005) 185-193.
96. Liu, S.J., Zhang, D.Q., Sui, X.M., Zhang, L., Cai, Z.W., Sun, L.Q., Liu, Y.J., Xue, Y. and Hu, G.F. The inhibition of *in vivo* tumorigenesis of osteosarcoma (OS)-732 cells by antisense human osteopontin RNA. **Cell. Mol. Biol. Lett.** 13 (2008) 11-19.
97. Miyamoto, T., Sato, H., Yogev, L., Kleiman, S., Namiki, M., Koh, E., Sakugawa, N., Hayashi, H., Ishikawa, M., Lamb, D.J. and Sengoku, K. Is a genetic defect in Fkbp6 a common cause of azoospermia in humans? **Cell. Mol. Biol. Lett.** 11 (2006) 557-569.
98. Wisniewska, A., Draus, J. and Subczynski, W.K. Is a fluid-mosaic model of biological membranes fully relevant? Studies on lipid organization in model and biological membranes. **Cell. Mol. Biol. Lett.** 8 (2003) 147-159.
99. Wladyka, B. and Pustelny, K. Regulation of bacterial protease activity. **Cell. Mol. Biol. Lett.** 13 (2008) 212-229.
100. Cottage, A., Mullan, L., Portela, M.B., Hellen, E., Carver, T., Patel, S., Vavouri, T., Elgar, G. and Edwards, Y.J. Molecular characterisation of the SAND protein family: a study based on comparative genomics, structural bioinformatics and phylogeny. **Cell. Mol. Biol. Lett.** 9 (2004) 739-753
101. Gronemeyer, H. and Miturski, R. Molecular mechanisms of retinoid action. **Cell. Mol. Biol. Lett.** 6 (2001) 3-52.
102. Agoston, V., Cemazar, M., Kajan, L. and Pongor, S. Graph-representation of oxidative folding pathways. **BMC Bioinformatics** 6 (2005) 19.
103. Kajan, L., Kertesz-Farkas, A., Franklin, D., Ivanova, N., Kocsor, A. and Pongor, S. Application of a simple likelihood ratio approximant to protein sequence classification. **Bioinformatics** 22 (2006) 2865-2869.
104. Kocsor, A., Kertesz-Farkas, A., Kajan, L. and Pongor, S. Application of compression-based distance measures to protein sequence classification: a methodological study. **Bioinformatics** 22 (2006) 407-412.
105. Vlahovicek, K., Kajan, L., Agoston, V. and Pongor, S. The SBASE domain sequence resource, release 12: prediction of protein domain-architecture using support vector machines. **Nucleic Acids Res.** 33 (2005) D223-D225.

106. Vlahovicek, K., Kajan, L., Murvai, J., Hegedus, Z. and Pongor, S. The SBASE domain sequence library, release 10: domain architecture prediction. **Nucleic Acids Res.** 31 (2003) 403-405.
107. von Grotthuss, M., Plewczynski, D., Ginalski, K., Rychlewski, L. and Shakhnovich, E.I. PDB-UF: database of predicted enzymatic functions for unannotated protein structures from structural genomics. **BMC Bioinformatics** 7 (2006) 53.
108. Wyrwicz, L.S., Koczyk, G., Rychlewski, L. and Plewczynski, D. ProteinSplit: splitting of multi-domain proteins using prediction of ordered and disordered regions in protein sequences for virtual structural genomics. **J. Phys. Condens. Matter** 19 (2007) 285222.
109. Grabarkiewicz, T., Grobelny, P., Hoffmann, M. and Mielcarek, J. DFT study on hydroxy acid-lactone interconversion of statins: The case of fluvastatin. **Org. Biomol. Chem.** 4 (2006) 4299-4306.
110. Grabarkiewicz, T. and Hoffmann, M. Syn- and anti-conformations of 5'-deoxy- and 5'-O-methyl-uridine 2',3'-cyclic monophosphate. **J. Mol. Model.** 12 (2006) 205-212.
111. Hoffmann, M., Chrzanowska, M., Hermann, T. and Rychlewski, J. Modeling of purine derivatives transport across cell membranes based on their partition coefficient determination and quantum chemical calculations. **J. Med. Chem.** 48 (2005) 4482-4486.
112. Hoffmann, M. and Marciniak, B. Quantum chemical study of the mechanism of ethylene elimination in silylative coupling of olefins. **J. Mol. Model.** 13 (2007) 477-483.
113. Hoffmann, M., Plutecka, A., Rychlewska, U., Kucybala, Z., Paczkowski, J. and Pyszka, I. New type of bonding formed from an overlap between pi aromatic and pi C=O molecular orbitals stabilizes the coexistence in one molecule of the ionic and neutral meso-ionic forms of imidazopyridine. **J. Phys. Chem. A Mol. Spectrosc. Kinet. Environ. Gen. Theory** 109 (2005) 4568-4574.
114. Hoffmann, M. and Rychlewski, J. Effects of substituting a OH group by a F atom in D-glucose. Ab initio and DFT analysis. **J. Am. Chem. Soc.** 123 (2001) 2308-2316.
115. Hoffmann, M., Rychlewski, J., Chrzanowska, M. and Hermann, T. Mechanism of activation of an immunosuppressive drug: azathioprine. Quantum chemical study on the reaction of azathioprine with cysteine. **J. Am. Chem. Soc.** 123 (2001) 6404-6409.
116. Plutecka, A., Hoffmann, M., Rychlewska, U., Kucybala, Z., Paczkowski, J. and Pyszka, I. Relationship between structure and photoinitiating abilities of selected bromide salts of 2-oxo-2,3-dihydro-1H-imidazo[1,2-a]pyridine (IMP): influence of the solvent and the substitution in benzaldehyde on the course of its reaction with IMP. **Acta Crystallogr. B** 62 (2006) 135-142.
117. Hoffmann, M., Eitner, K., von Grotthuss, M., Rychlewski, L., Banachowicz, E., Grabarkiewicz, T., Szkoda, T. and Kolinski, A. Three

- dimensional model of severe acute respiratory syndrome coronavirus helicase ATPase catalytic domain and molecular design of severe acute respiratory syndrome coronavirus helicase inhibitors. **J. Comput. Aided Mol. Des.** 20 (2006) 305-319.
118. Ostrowski, J., Rubel, T., Wyrwicz, L.S., Mikula, M., Bielasik, A., Butruk, E. and Regula, J. Three clinical variants of gastroesophageal reflux disease form two distinct gene expression signatures. **J. Mol. Med.** 84 (2006) 872-882.
119. Paziewska, A., Wyrwicz, L.S., Bujnicki, J.M., Bomsztyk, K. and Ostrowski, J. Cooperative binding of the hnRNP K three KH domains to mRNA targets. **FEBS Lett.** 577 (2004) 134-140.
120. Paziewska, A., Wyrwicz, L.S. and Ostrowski, J. The binding activity of yeast RNAs to yeast Hek2p and mammalian hnRNP K proteins, determined using the three-hybrid system. **Cell. Mol. Biol. Lett.** 10 (2005) 227-235.
121. von Grothuss, M., Koczyk, G., Pas, J., Wyrwicz, L.S. and Rychlewski, L. Ligand.Info small-molecule Meta-Database. **Comb. Chem. High. Throughput Screen.** 7 (2004) 757-761.
122. von Grothuss, M., Pas, J. and Rychlewski, L. Ligand-Info, searching for similar small compounds using index profiles. **Bioinformatics** 19 (2003) 1041-1042.
123. von Grothuss, M., Wyrwicz, L.S. and Rychlewski, L. mRNA cap-1 methyltransferase in the SARS genome. **Cell** 113 (2003) 701-702.
124. Wyrwicz, L.S. and Rychlewski, L. Herpes glycoprotein gL is distantly related to chemokine receptor ligands. **Antiviral Res.** 75 (2007) 83-86.
125. Zemojtel, T., Frohlich, A., Palmieri, M.C., Kolanczyk, M., Mikula, I., Wyrwicz, L.S., Wanker, E.E., Mundlos, S., Vingron, M., Martasek, P. and Durner, J. Plant nitric oxide synthase: a never-ending story? **Trends Plant Sci.** 11 (2006) 524-525; author reply 526-528.
126. Plewczynski, D., Hoffmann, M., von Grothuss, M., Ginalski, K. and Rychlewski, L. In silico prediction of SARS protease inhibitors by virtual high throughput screening. **Chem. Biol. Drug Design** 69 (2007) 269-279.
127. Plewczynski, D., Hoffmann, M., von Grothuss, M., Knizewski, L., Rychlewski, L., Eitner, K. and Ginalski, K. Modelling of potentially promising SARS protease inhibitors. **J. Phys. Condens. Matter** 19 (2007) 285207.
128. Feder, M., Pas, J., Wyrwicz, L.S. and Bujnicki, J.M. Molecular phylogenetics of the RrmJ/fibrillarlin superfamily of ribose 2'-O-methyltransferases. **Gene** 302 (2003) 129-138.
129. Ginalski, K., Pas, J., Wyrwicz, L.S., von Grothuss, M., Bujnicki, J.M. and Rychlewski, L. ORFeus: Detection of distant homology using sequence profiles and predicted secondary structure. **Nucleic Acids Res.** 31 (2003) 3804-3807.

130. Klimek-Tomczak, K., Mikula, M., Dzwonek, A., Paziewska, A., Wyrwicz, L.S., Hennig, E.E. and Ostrowski, J. Mitochondria-associated satellite I RNA binds to hnRNP K protein. **Acta Biochim. Pol.** 53 (2006) 169-178.
131. Klimek-Tomczak, K., Wyrwicz, L.S., Jain, S., Bomsztyk, K. and Ostrowski, J. Characterization of hnRNP K protein-RNA interactions. **J. Mol. Biol.** 342 (2004) 1131-1141.
132. Pas, J., von Grotthuss, M., Wyrwicz, L.S., Rychlewski, L. and Barciszewski, J. Structure prediction, evolution and ligand interaction of CHASE domain. **FEBS Lett.** 576 (2004) 287-290.
133. von Grotthuss, M., Pas, J., Wyrwicz, L., Ginalski, K. and Rychlewski, L. Application of 3D-Jury, GRDB, and Verify3D in fold recognition. **Proteins** 53 Suppl 6 (2003) 418-423.
134. von Grotthuss, M., Plewczynski, D., Ginalski, K., Rychlewski, L. and Shakhnovich, E.I. PDB-UF: database of predicted enzymatic functions for unannotated protein structures from structural genomics. **BMC Bioinformatics** 7 (2006) 53.
135. von Grotthuss, M., Wyrwicz, L.S., Pas, J. and Rychlewski, L. Predicting protein structures accurately. **Science** 304 (2004) 1597-1599; author reply 1597-1599.
136. Wyrwicz, L.S., von Grotthuss, M., Pas, J. and Rychlewski, L. How unique is the rice transcriptome? **Science** 303 (2004) 168; author reply 168.
137. Plewczynski, D., Jaroszewski, L., Godzik, A., Kloczkowski, A. and Rychlewski, L. Molecular modeling of phosphorylation sites in proteins using a database of local structure segments. **J. Mol. Mod.** 11 (2005) 431-438.
138. Plewczynski, D., Tkacz, A., Godzik, A. and Rychlewski, L. A support vector machine approach to the identification of phosphorylation sites. **Cell. Mol. Biol. Lett.** 10 (2005) 73-89.
139. Plewczynski, D., Tkacz, A., Wyrwicz, L., Godzik, A., Kloczkowski, A. and Rychlewski, L. Support-vector-machine classification of linear functional motifs in proteins. **J. Mol. Mod.** 12 (2006) 453-461.
140. Plewczynski, D., Tkacz, A., Wyrwicz, L.S. and Rychlewski, L. AutoMotif server: prediction of single residue post-translational modifications in proteins. **Bioinformatics** 21 (2005) 2525-2527.
141. Plewczynski, D., Tkacz, A., Wyrwicz, L.S., Rychlewski, L. and Ginalski, K. AutoMotif Server for prediction of phosphorylation sites in proteins using support vector machine: 2007 update. **J. Mol. Mod.** 14 (2008) 69-76.
142. Plewczynski, D., Slabinski, L., Tkacz, A., Kajan, L., Holm, L., Ginalski, K. and Rychlewski, L. The RPSP: Web server for prediction of signal peptides. **Polymer** 48 (2007) 5493-5496.
143. Fernandez-Ballester, G. and Serrano, L. Prediction of protein-protein interaction based on structure. **Methods Mol. Biol.** 340 (2006) 207-234.
144. Plewczynski, D., Pas, J., von Grotthuss, M. and Rychlewski, L. Comparison of proteins based on segments structural similarity). **Acta Bioch. Pol.** 51 (2004) 161-172.

145. Plewczynski, D., Rychlewski, L., Ye, Y.Z., Jaroszewski, L. and Godzik, A. Integrated web service for improving alignment quality based on segments comparison. **BMC Bioinformatics** 5 (2004) 98.
146. Kinch, L.N., Ginalski, K., Rychlewski, L. and Grishin, N.V. Identification of novel restriction endonuclease-like fold families among hypothetical proteins. **Nucleic Acids Res.** 33 (2005) 3598-3605.
147. Ginalski, K., Elofsson, A., Fischer, D. and Rychlewski, L. 3D-Jury: a simple approach to improve protein structure predictions. **Bioinformatics** 19 (2003) 1015-1018.
148. Ginalski, K. and Rychlewski, L. Detection of reliable and unexpected protein fold predictions using 3D-Jury. **Nucleic Acids Res.** 31 (2003) 3291-3292.
149. Bu, D., Zhao, Y., Cai, L., Xue, H., Zhu, X., Lu, H., Zhang, J., Sun, S., Ling, L., Zhang, N., Li, G. and Chen, R. Topological structure analysis of the protein-protein interaction network in budding yeast. **Nucleic Acids Res.** 31 (2003) 2443-2450.
150. Sen, T.Z., Kloczkowski, A. and Jernigan, R.L. Functional clustering of yeast proteins from the protein-protein interaction network. **BMC Bioinformatics** 7 (2006) 355.
151. Ogmen, U., Keskin, O., Aytuna, A.S., Nussinov, R. and Gursoy, A. PRISM: protein interactions by structural matching. **Nucleic Acids Res.** 33 (2005) W331-W336.
152. Aytuna, A.S., Gursoy, A. and Keskin, O. Prediction of protein-protein interactions by combining structure and sequence conservation in protein interfaces. **Bioinformatics** 21 (2005) 2850-2855.
153. Aloy, P., Bottcher, B., Ceulemans, H., Leutwein, C., Mellwig, C., Fischer, S., Gavin, A.C., Bork, P., Superti-Furga, G., Serrano, L. and Russell, R.B. Structure-based assembly of protein complexes in yeast. **Science** 303 (2004) 2026-2029.
154. Aloy, P. and Russell, R.B. Interrogating protein interaction networks through structural biology. **Proc. Natl. Acad. Sci. USA** 99 (2002) 5896-5901.
155. Aloy, P. and Russell, R.B. InterPreTS: protein interaction prediction through tertiary structure. **Bioinformatics** 19 (2003) 161-162.
156. Ben-Hur, A. and Noble, W.S. Kernel methods for predicting protein-protein interactions. **Bioinformatics** 21 Suppl 1 (2005) i38-46.
157. Ben-Hur, A. and Noble, W.S. Choosing negative examples for the prediction of protein-protein interactions. **BMC Bioinformatics** 7 Suppl 1 (2006) S2.
158. Gomez, S.M., Noble, W.S. and Rzhetsky, A. Learning to predict protein-protein interactions from protein sequences. **Bioinformatics** 19 (2003) 1875-1881.
159. Nanni, L. and Lumini, A. An ensemble of K-local hyperplanes for predicting protein-protein interactions. **Bioinformatics** 22 (2006) 1207-1210.

160. Sun, S., Zhao, Y., Jiao, Y., Yin, Y., Cai, L., Zhang, Y., Lu, H., Chen, R. and Bu, D. Faster and more accurate global protein function assignment from protein interaction networks using the MFGO algorithm. **FEBS Lett.** 580 (2006) 1891-1896.
161. Bordner, A.J. and Abagyan, R.A. Large-scale prediction of protein geometry and stability changes for arbitrary single point mutations. **Proteins** 57 (2004) 400-413.
162. Lu, H., Zhu, X., Liu, H., Skogerbo, G., Zhang, J., Zhang, Y., Cai, L., Zhao, Y., Sun, S., Xu, J., Bu, D. and Chen, R. The interactome as a tree-an attempt to visualize the protein-protein interaction network in yeast. **Nucleic Acids Res.** 32 (2004) 4804-4811.
163. Plewczynski, D., Spieser, S.A.H. and Koch, U. Assessing different classification methods for virtual screening. **J. Chem. Inf. Mod.** 46 (2006) 1098-1106.
164. Plewczynski, D., von Grotthuss, M., Spieser, S.A.H., Rychlewski, L., Wyrwicz, L.S., Ginalski, K. and Koch, U. Target specific compound identification using a support vector machine. **Comb. Chem. High Throughput Screen.** 10 (2007) 189-196.
165. Plewczynski, D., Spieser, S.A. and Koch, U. Assessing different classification methods for virtual screening. **J. Chem. Inf. Model.** 46 (2006) 1098-1106.
166. Sen, T.Z., Kloczkowski, A., Jernigan, R.L., Yan, C., Honavar, V., Ho, K.M., Wang, C.Z., Ihm, Y., Cao, H., Gu, X. and Dobbs, D. Predicting binding sites of hydrolase-inhibitor complexes by combining several methods. **BMC Bioinformatics** 5 (2004) 205.
167. Donald, J.E., Hubner, I.A., Rotemberg, V.M., Shakhnovich, E.I. and Mirny, L.A. CoC: a database of universally conserved residues in protein folds. **Bioinformatics** 21 (2005) 2539-2540.
168. Mirny, L.A., Abkevich, V.I. and Shakhnovich, E.I. How evolution makes proteins fold quickly. **Proc. Natl. Acad. Sci. USA** 95 (1998) 4976-4981.
169. Mirny, L.A. and Shakhnovich, E.I. Universally conserved positions in protein folds: reading evolutionary signals about stability, folding kinetics and function. **J. Mol. Biol.** 291 (1999) 177-196.