Review Article

Virtual Interactomics of Proteins from Biochemical Standpoint

Jaroslav Kubrycht,[1] Karel Sigler,[2] and Pavel Souček[3]

[1] *Department of Physiology, Second Medical School, Charles University, 150 00 Prague, Czech Republic*
[2] *Laboratory of Cell Biology, Institute of Microbiology, Academy of Sciences of the Czech Republic, 142 20 Prague, Czech Republic*
[3] *Toxicogenomics Unit, National Institute of Public Health, 100 42 Prague, Czech Republic*

Correspondence should be addressed to Jaroslav Kubrycht, jkub@post.cz

Virtual interactomics represents a rapidly developing scientific area on the boundary line of bioinformatics and interactomics. Protein-related virtual interactomics then comprises instrumental tools for prediction, simulation, and networking of the majority of interactions important for structural and individual reproduction, differentiation, recognition, signaling, regulation, and metabolic pathways of cells and organisms. Here, we describe the main areas of virtual protein interactomics, that is, structurally based comparative analysis and prediction of functionally important interacting sites, mimotope-assisted and combined epitope prediction, molecular (protein) docking studies, and investigation of protein interaction networks. Detailed information about some interesting methodological approaches and online accessible programs or databases is displayed in our tables. Considerable part of the text deals with the searches for common conserved or functionally convergent protein regions and subgraphs of conserved interaction networks, new outstanding trends and clinically interesting results. In agreement with the presented data and relationships, virtual interactomic tools improve our scientific knowledge, help us to formulate working hypotheses, and they frequently also mediate variously important *in silico* simulations.

1. General Remarks

Many important findings in pharmacology, cell biology, and pathobiology have been achieved with the aid of virtual interactomics including computer-aided structural analysis, prediction and *in silico* simulation of interacting sites, protein complexes, and interaction networks. Virtual interactomics has been developed in the last thirty years, and it is in fact based on gradual bioinformatic processing of experimental data. These data were usually obtained from individual studies of interactions, and various large-scale experimental methods such as the two-hybrid system, phage display library studies reverse interactomics, SPOT arrays or microarray studies, and extended sequence studies [1–7].

In addition to sequence data, three-dimensional (3D) structures are ever more frequently required for interactomic predictions. X-ray crystallography or nuclear magnetic resonance studies represent the most frequent sources of 3D

structures, whereas combination of electron microscopy of molecular complexes with X-ray crystallography turns out to be interesting for the same purpose [8–11]. Alternatively, sophisticated 3D structure simulations such as homology modeling or combination of cryoelectron microscopy densities, and molecular dynamics appear to be also useful for approximating conventional 3D input at least in some cases [12–14]. In addition to 3D shape, solvent and surface accessibilities (or more likely actual dynamic accessibility following from accompanying interacting structures or proteolysis; cf. [15, 16]) were considered to be important criteria for reevaluation of possible interaction sites. Many experimentally investigated and predicted structural relationships were also stored in interactomic databases to be selectively found, processed and compared. Moreover, some interaction data differently stored in multiple databases have been searched with the aid of special data mining servers such as Dasmiweb and PINA v2.0 ([17, 18]; see also Table 1).

TABLE 1: Some more recent online accessible bioinformatic tools.

Programs	Purpose	Input	Evaluation tools and procedures	Compared structures	Internet accessibility	References
Program tools						
3D-BLAST (two methods)	*Identification of 23 states of the structural alphabet; *Comparison of protein folds using spherical polar Fourier basis functions.	PDB ACF	*SASM; *Carbo-like similarity score	*structural alphabet sequence databases; *SPF shape density rep.	http://3d-blast.life.nctu.edu.tw/ ; http://threedblast.loria.fr/	[41, 42]
DASMIweb	Online integration, analysis and assessment of distributed protein interaction data and predictions (interactomic mining)	some protein identifiers	literature curation, prediction, 3D analysis	records of interactions	http://www.dasmiweb.de	[17]
FFAS03	Server accepts a user supplied protein sequence and automatically generates a profile, which is then compared with several sets of sequence profiles of proteins databases	QS	PPA	DS	http://ffas.godziklab.org/	[36, 37]
KinasePhos 2.0	Prediction of kinase-specific phosphorylation sites based on the site sequences from databases	QS	k-fold + Jackknife cross-validations	phosphorylation sites from Phospho.ELM Swiss-Prot	http://KinasePhos2.mbc.nctu.edu.tw/	[62]
Phos3D	Method of phosphorylation-site prediction based on 3D structural information associated with 530 phys-chem-pr	QS (+3D context)	SVM, spatial amino acid propensities	PDB coordinates	http://phos3d.mpimp-golm.mpg.de/	[63]
SeSAW	Identification of functionally or evolutionarily conserved motifs based on balancing between sequence and structural similarities	*PDB ID; *PDB QF; ID, IDt	P value SeSAW score	conserved domains templates	http://sysimm.ifrec.osaka-u.ac.jp/SeSAW/	[61]
Databases						
ADAN	Prediction of protein-protein interactions of different modular protein domains mediated by linear motifs	PDB ACF	PSSM_PI	integrated DaSt	http://adan-embl.ibmc.umh.es/	[64]
CDD (RPS BLAST)	Conserved domain relationships, domain location of frequent binding sites	QS	model MSA derived PSSM	Conserved domains (PSSM)	http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml ; http://www.ncbi.nlm.nih.gov/BLAST	[65–67]
GWIDD	Docking database; integrated resource for studies of protein-protein interactions on the genome scale	search interface	docking techniques	DaSt	http://gwidd.bioinformatics.ku.edu	[68]
MegaMotifBase (structural database)	3D motif orientation, intermotif distances, solvent accessibility, secondary structure content, hydrogen bonding, residual packaging, familiar/superfamiliar/none motif relationship	motif or sequence pattern	complex evaluation including projection	DS + DaSt	http://caps.ncbs.res.in/MegaMotifBase/index.html	[69]

TABLE 1: Continued.

Programs	Purpose	Input	Evaluation tools and procedures	Compared structures	Internet accessibility	References
PTGL	Database for secondary structure-based protein topologies; visualization of topology diagrams and 3D structures	PDB QF	ULN4D-graph theory	DS + DaSt	http://ptgl.uni-frankfurt.de/	[70]
RsiteDB	Database of protein binding pockets which interact with single strand RNA	PDB code, UPS	3D arrangement of phys-chem-pr	3D-CBP	http://bioinfo3d.cs.tau.ac.il/RsiteDB/	[71]

*Independent alternatives; 3D-CBP: 3D consensus binding patterns important for protein–nucleotide recognition; DS: database sequences; DSA: double sequence alignment; DaSt: database structures; CS: list of compared sequences; ID, IDt: chain or template ID, respectively; MSA: multiple sequence alignment; QS: query sequence(s) (instead of QS, clonal names or gi numbers can be used in majority of given approaches); phys-chem-pr: physicochemical properties; PDB: Protein data bank; PDB QF: PDB-formatted query file; PDB ACF: PDB-derived atom coordinate file; PPA: profile–profile alignment; PSSM: position-specific scoring matrices; PSSM_PI: PSSM for protein-protein interactions calculated by FoldX; rep.: representations; SASM: structural alphabet substitution matrix; SPF: spherical polar Fourier; SVM: support vector machines; ULN4D-graph theory: unique linear notations of four descriptions for protein structures on different abstraction levels based on graph theory; UPS: unbound protein structure.

Virtual interactomic studies are mostly realized via computer programs including numerous online accessible tools. Three types of computer data processing are important for online interactomic predictions on molecular level, that is, (i) structure comparisons, (ii) molecular docking studies, and (iii) reevaluation of current database data, and accessible or proposed protein interaction networks. Similarly, three types of interacting structures can be distinguished with respect to their different molecular origin. This concerns (i) conserved structures, (ii) randomly/quasi-randomly *in vitro* generated or rapidly evolved structures (e.g., mimotopes and disordered regions), and (iii) binding sites of antigen receptors expressed in specific immune cell clones, which are usually developed during the regular recombination process, and later also at the time of immune response.

In contrast to reactomics, interactomics described here deals exclusively with interactions, and thus concerns only interactions of enzyme active sites but not their subsequent reaction mechanism. Consequently, modeling of enzyme reactions exceeds the topic of this review. In addition, since we dealt here with protein interactomics, this paper also does not contain information about interactions of DNA with nonpeptide ligands.

2. Structural Similarities of Interacting Sites

By using various structurally based programs, many protein interactions can be predicted based on the occurrence of phylogenetically conserved or convergently developed functionally important common structural features (motifs, sequence patterns, consensi, constructs of conserved domain sequences, supersecondary structures, supersecondary motifs, 3D-arranged structural patterns and pockets). However, diversification within protein families, and superfamilies causes losses of interacting structures or disables their accessibility. On the other hand, new interactive pairs frequently appear in cases of disordered protein regions, that is, peptide segments naturally occurring in multiple conformation variants [19–21]. The attendant stability problems, as well as some additional problems with molecular analysis, were diminished by the development of databases enabling reevaluation of selected structural relationships. The databases inform us about similar or common structural features, frequent locations of binding sites in related domains, solvent accessibility, and location of investigated segments in 3D structures of proteins (see Table 1).

During the last twenty years, conventional sequence-based search for conserved structures frequently combined different evaluations of sequence similarities. The corresponding protocols usually combined double sequence, and multiple sequence comparisons like BLASTP, PSI-BLAST, RPS-BLAST, Clustal W and MUSCLE [22, 23]. In addition, highly selective PHI-BLAST or PSI-BLAST searches with specifically restricted representative query sequences such as consensi (and also sequence patterns) made it possible to locate the corresponding potentially interacting sites

in extended sequence sets [23–26]. Except for current conserved sequences, extremely variable but still defined structures such as heptad repeats were investigated for purposes of molecular topology [27, 28]. These repeats can be written as an alphabet with generalized characters representing aa groups instead of individual aa (hydrophobic, charged, polar, etc.). Together with usual (motif-related sequence block derived) patterns an important part of heptad repeats can be searched in database sequences and partially reevaluated by means of PROSITE programs [29–32]. Apart from regions with highly conserved sequences, additional conserved structures were found when combining evaluation of primary and secondary structures [33], PSI-BLAST and secondary structure [34], or when using fold recognition [34, 35]. The compared sequence queries were moreover evaluated on the sophisticated widely used FFAS03 server (about 250 references) providing the third generation of the profile-profile alignment and fold-recognition algorithm mediated by program FFAS (fold, and function assignment system; [36, 37]; Table 1). The sensitivity FFAS-related profile-profile comparison is now widely recognized and many Web servers implementing such algorithms are available, for example, HHPRED, COMPASS, COMA, PHYRE, GenThreader, FORTE and webPRC [37]. More recent multiple sequence alignment program BCL::Align includes also combined evaluation of structural similarities [38] (for applications, see, e.g., [39, 40]). The corresponding scoring function is a weighted sum of scores derived based on (i) the traditional PAM, and BLOSUM scoring matrices, (ii) position-specific scoring matrices by PSI-BLAST, (iii) secondary structure predicted by a variety of methods, (iv) chemical properties, and (v) gap penalties. Monte Carlo algorithm was then used to determine necessary optimized weights in cases of sequence alignment and fold recognition.

Input of 3D coordinates or their transformed representations was necessary for other structural studies predicting also functional interaction sites. The corresponding research yielded two servers with different 3D-BLAST programs enabling us to compare folds and fold families [41, 42] (for details see also Table 1). Alternative structural comparison substituting 3D relationships was performed when searching for the maximum contact map overlaps [43]. The extended contact map comparison appeared very early [44]. Contact map is in fact determined by the matrix of distances between individual amino acids, contact threshold and specification of contact types [45, 46]. This map can be visualized by CMView software [46]. Similarly to sequence motifs, conserved patterns of 3D peptide arrangements (CP-3D-A) were considered as an additional type of structure-function motifs. These structures were recorded by sequence independent 3D-templates and can be searched on Evolutionary Trace Annotation server [47, 48]. Frequent functionally important CP-3D-A occurred, for instance, in active sites of proteins containing porphyrin rings including members of cytochrome P450 superfamily [49]. Another recent interesting approach consisted in generalized motif search in large alphabet (cumulative occurrence of several alphabets in this case) inputs including simultaneously evaluation of DNA sequence, protein sequence and supersecondary structure

motifs [50]. According to the authors, large alphabets are important in cases when structures share little similarity at primary level.

Similarly to multiple sequence alignment, various forms of multiple structural alignments (MSTA) have been generated. Older attempts at MSTA were based on secondary structural alignment [51–53]. On the other hand, recent MSTA approaches determined common spatial (3D) structures. These approaches include new strategies employing "molecular sieving" of protein structures, minimization of an energy function over the low-dimensional space of the relative rotations, and translations of the molecules, geometric hashing and contact-window-derived motif library [54–58].

In spite of the increasing possibility of database reevaluation, some structurally based programs predicting interacting sites became at least relatively autonomous. A large number of specific and relatively autonomous programs concerned the prediction of epitopes (for details see Section 3). Similarly, various possibilities of prediction of phosphorylation sites were frequently investigated (two examples in Table 1), since phosphorylation reactions represent an important signaling and regulatory network in cell biology and pathobiology. The interest followed also from the extended building of databases related to phosphorylation sites, which brought many interesting insights. For instance, in accordance with the linear motif atlas for phosphorylation-dependent signaling, tyrosine kinases mutated in cancer exhibited lower specificity than their non-oncogenic relatives [59]. Similarly, collection and motif-based prediction of phosphorylation sites in human virus proteins suggested a substantial role for human kinases in regulation/mediation of viral protein functions [60]. In contrast to programs predicting specific types of interacting sites, SeSAW represents an example of general online accessible program allowing prediction of possible functionally important structures [61] (see also Table 1). The corresponding balance between different combined structural evaluations concerned, among others, data present in position specific scoring matrix (PSSM), template-derived PSSM and template functional annotations.

3. Mimotope and Epitope Interactions

Epitopes are defined as the structures responsible for interaction of antigens with binding sites of antigen receptors. On the other hand, mimotopes belong to artificially prepared peptides, which interact with natural templates (most likely proteins), and thus mimic other peptides or organic compounds in their functionally important interactions. Mimotope development is based on synthetic peptide or phage display libraries, whereas natural development of specific cell clones is necessary for epitope recognition by specific antigen receptors. This means that both epitopes and mimotopes can sometimes considerably differ from the usual conserved structures mentioned above. In spite of the described difference, a unifying point between mimotopes and epitopes exists, because mimotopes were originally defined as peptides mimicking epitopes [72], forming thus

only a subset of the later current mimotope repertoire. In addition to this historical linkage, rapidly developing (diverging) structures such as molecular mimicry enabling parasitic attack and adaptation of pathogenic viruses and bacteria, disordered regions of proteins, and protein loops appear to be good candidates for extended investigation of mimotope similarities (cf. [21, 73–75]) and mimotope based prediction (see below).

Mimotopes were originally derived in studies with a phage display library. This phage technology was discovered in the eighties [76]. The corresponding boom in the nineties then comprised novel random, partially randomized or gene-fragment-derived oligopeptide libraries able to functionally mimic epitopes, autoepitopes, short peptide ligands, protein kinase or proteinase substrates, as well as peptides mimicking organic substances such as biotin when interacting with steptavidin [77–85]. Mimotope similarities were frequently defined using sequence patterns, whereas additional types of nonsequence structural similarities were also described (see, e.g., [81, 83]). More recent biotechnological research of mimotopes yielded potential peptide drugs [86–89], peptide vaccines [90–92], and peptides suitable for specific (mostly nanoparticle mediated) drug delivery to tumor cells, brain, atherosclerotic plaques, and other therapeutically important sites of human or animal organisms [93–97]. In spite of this considerable progress, virtual interactomic tools are not still able to compare or predict organic drugs based on their effective spatial similarity with functionally active mimotopes.

Database registration and authentication of mimotopes have been performed for more than ten years [98–100]. Special programs comparing primary structures of epitopes were simultaneously developed (programs FINDMAP, and EPIMAP [101, 102]). Some of them employed also multiple-sequence alignment evaluation (program MIMOP [103]). Similarly, coexisting peptide databases were established to process the accumulated information. These databases (i) recognized sequence subsets classified after *in vitro* evolution of phage display libraries, (ii) offered many integrating programs, and (iii) made it possible to find all mimotope sets that have the 3D structure of a target-template complex (databases ASPD, RELIC and MimoDB [98, 99, 104, 105]). In addition, novel mimotope-assisted computer-aided epitope prediction was discovered. This prediction came from both the 3D structure of an interacting partner and sequences of similarly interacting mimotopes, and has also concerned some interacting partners different from specific antigen receptors and antigens (programs PepSurf, Pep-3D-Search, and MimoPro [106–108]). Based on this approach, improved specificity, and extended the repertoire of predicted epitopes or other interacting partners were achieved. Further progress in programming then resulted in accelerated computation in spite of more complicated, and precise strategies of data processing. In addition, pattern recognition algorithm was developed, which can effectively be employed to screen a mixture of antibodies, and define the breadth of epitopes recognized by polyserum directed against specific proteins [109]. This possibility appears to be interesting with respect to future vaccine design.

The recent status of epitope prediction is still far from resolved due to insufficient extent of the datasets, and still requires continuous improvement of database organization [110, 111]. In such state, mimotope-assisted epitope prediction mentioned above and combined approaches represent a certain improvement in the quality of epitope prediction. Online accessible combined approaches of B-cell epitope prediction mostly evaluate 3D structures together with solvent accessibility (programs CEP, ElliPro, PEPOP and 3D alternative of Epitopia [112–115]; Table 2). Similarly to many other combined approaches, combined restriction diminishes the number of false positivities but simultaneously can cause increased number of false negativities. For instance, the widely used requirement of solvent accessibility mentioned above would in fact eliminate at least part of conserved autoepitopes, which contain hydrophobic patterns (cf. [116–118]). To diminish losses following from employed combinations or effects of too strict (sure) thresholds, a metaserver was developed that sums up the results from six epitope-predicting servers [119]. In addition, some new types of online accessible prediction of linear epitopes appeared to complete the preceding results [120, 121].

An interesting input simplification has been achieved with a novel web server CBTOBE. This server uses learning of support vector machines based on physicochemical profiles, and makes it possible to predict conformation epitopes based on a sole sequence input [122]. A recent private alternative of CBTOBE was based on older evaluation of secondary structure and solvent accessibility (server COBEpro [123]) further complemented by evolutionary information and machine-learning-derived evaluation [124].

Combined predictions of epitopes presented to T-cell receptors have integrated class I MHC (major histocompatibility complex) peptide binding affinity, TAP transport efficiency and prediction of proteasome cleavage (programs EpiJen, NetCTL-1.2, FRED [125–128]). Nevertheless, independent simulations of peptide-binding affinity of various MHC molecules have been also proposed [129] as well as the corresponding neural network-based learning [130]. Both combined, and sole approaches then represent starting steps for further reevaluation with respect to T-cell receptor interactions, for example, using learning of support vector machines, and strict kernels based on 531 physicochemical properties (POPISK [131]). Important information about existing T-cell, and B-cell epitopes can be also obtained from the Immune Epitope Database (IEDB [132, 133]). Among others, EpitopeViewer of the 3D structural subcomponent of IEDB (IEDB-3D) allows the user to visualize, render and analyze the structure, and save structural and contact views as high-quality pictures for publication [133].

Since production of various vaccines, and detection kits with mononoclonal antibodies requires high efficiency of preparations, special searches for conserved epitopes have been developed for this purpose. Though the meaning of crucial term "conserved epitope" exceeds primary structure relationships, the repertoires of the corresponding structures appear to be sometimes considerably limited. In fact, structural or regulatory adaptations of pathogenic microogranisms and viruses cause less stability of

immunologically important, and promising promiscuous (cytotoxic or helper) T-cell epitopes broadly cross-reacting with different MHC antigens, different frequencies of T-cell epitopes specific usually only for certain unique MHC molecule, as well as losses of immunogenicity or even absence of immune response. In addition to current evaluation, drug effects (cf. effects in docking studies described below) and accompanying structural variability (e.g., enzyme polymorphisms or special pathogenetical effects) have to be considered with respect to possible peptide epitope modifications, possible bias, or improvement of therapeutical design. It is a question whether conserved spatial structures following from multiple structural alignments mentioned above can be also interesting for prediction of conserved conformation of B-cell epitopes. The second part of Table 2 contains some web servers interesting with respect to prediction of conserved epitopes, whereas selected examples related to AIDS research follow.

The first more complex prediction of HIV epitopes was based on sequence conservativeness, secondary structure, solvent accessibility, hydrophilicity and flexibility [134]. The study pointed to an unfavorably frequent occurrence of changes in secondary structures predicted as antigenic. Certain progress in the research of conserved HIV-1 epitopes was achieved when assessing their possible interactions with MHC antigens. This comprised peptide prediction based on EpiMatrix score [135], construction of the peptide property model from a training dataset [136, 137] and combined T-cell epitope evaluation mentioned above [126, 127]. Lately, phylogenetic hidden Markov models allowed to predict HIV-1-related T-cell epitopes based on contiguous aa positions that evolve under immune pressure dependent among others on host HLA alleles [138]. In a recent paper, structure-function analysis based on a specifically devised mathematical model revealed that protection from neutralization (shielding of neutralization-sensitive domains) is enforced by intersubunit contact between the variable loops 1 and 2 (gp120 V1V2) of HIV-1, and domains of neighboring gp120 subunits in the trimer encompassing the V3 loop [139].

4. Protein Docking

Molecular docking is a method, which predicts the preferred reciprocal orientation of two molecules when they bound to each other to form a stable complex [147]. In case of simulated protein interactions, the authors currently speak about protein docking rather than about molecular docking, since molecular docking represents a term comprising also nucleic acid interactions with nonpeptide molecules. Formerly, molecular docking simulated "*lock-and-key*" type of protein interactions. Its original variants appeared in the eighties and reassumed interpretations of older molecular graphics [148–152]. These approaches were restricted by complementarity demands or simplified requirements for energy minimization. Lately, "*hand-in-glove*" analogy was found to be more appropriate for molecular docking than the "*lock-and-key*" one [153]. In addition to this conventional model, three new models have been recently developed for

TABLE 2: Epitope prediction or reevaluation on accessible servers.

Programs	Purpose	Tools or procedures of evaluation	Input	Output	Internet accessibility	References
				Epitope prediction		
MimoDB 2.0	Mimotope database	MySQL relational database	according to menu	Structures visualization, alignments, and so forth	http://immunet.cn/mimodb/	[105]
MimoPro	Maps a group of mimotopes back to a source antigen so as to locate the interacting epitope on the antigen	Branch and bound optimization (analysis of overlapping patches on the surface of a protein)	PDB identifier mimotope sequences	Score + 3D location in antigen	http://informatics.nenu.edu.cn/MimoPro/	[108]
ViPR	Virus pathogen database	Integration of various resources	according to menu	Structures, annotations, and so forth	http://www.viprbrc.org/	[140]
MetaMHC	Prediction of MHC binding epitopes	meta-approach	QS	Four metapredictor scores including MetaSVMp score	http://www.biokdd.fudan.edu.cn/Service/MetaMHC.html	[141]
Epitopia	Prediction of B-cell epitopes	*MLA *MLA + 3D +SA	*QS *PDB ACF	Immunogenicity score + probability score + color scale record on QS	http://epitopia.tau.ac.il	[115]
ElliPro	New structure-based tool for the prediction of antibody epitopes	3D + SA + flx + antigenicity	*QS *PDB ACF	Score + visualized epitope 3D structure and 3D location	http://tools.immuneepitope.org/tools/ElliPro	[114]
MHCPred	Prediction of class II mouse MHC peptide binding affinity	ISC-PLS, SYBYL software package	QS	Binding affinity (pIC50)	http://www.jenner.ac.uk/MHCPred	[129]
				Stable epitopes and vaccines		
BayesB	SVM-based prediction of linear B-cell epitopes using Bayes Feature Extraction	Residue conservation+ position-specific residue propensities	QS	Residue epitope propensity score	http://www.immunopred.org/bayesb/index.html	[120]
OptiTope	Selection of optimal peptides for epitope-based vaccines	Multistep parallel evaluation	MSA of Ag list of ES table: ES, ES related MHC	Fraction of overall immunogenicity covered MHC	http://www.epitoolkit.org/optitope	[142]
PVS	PVS returns a variability-masked sequence, which can be submitted to the RANKPEP server to predict conserved T-cell epitopes.	3D visualization of MSA derived sequence variability "per site"	PDB ACF	3D map of variability fragments with no variable residues and their 3D location in antigen	http://imed.med.ucm.es/PVS/	[143]

Table 2: Continued.

Programs	Purpose	Tools or procedures of evaluation	Input	Output	Internet accessibility	References
PEPVAC	Prediction of MHC I; server can also identify conserved and promiscuous MHC I ligands	PSSM, distance matrix, phylogenic clustering algorithm	Genome, HLA-supertypes	Selected peptide sequences score	http://immunax.dfci.harvard.edu/PEPVAC/	[144]
RANKPEP	Prediction of MHC I and MHC II ligands	Profile comparison	Genome, HLA-supertypes, QS	Selected peptide sequences score	http://www.mifoundation.org/Tools/rankpep.html	[145, 146]

*Independent alternatives; Ag: antigen; BFE: Bayes Feature Extraction; flx: flexibility; ES: epitope sequences; HLA: human leukocyte antigens (human MHC); ISC-PLS: iterative self-consistent partial-least-squares based additive method; MHC: alleles of major histocompatibility antigens; MLA: machine learning algorithm; MSA: multiple sequence alignment; PDB: protein data bank; PDB ACF: PDB-derived atom coordinate file; QS: query sequence(s); SA: solvent accessibility; SVM: support vector machines.

docking studies of protein-protein interactions [154], that is, (i) conformer selection model, using a novel ensemble docking algorithm, (ii) induced fit model, employing energy-gradient-based backbone minimization, and (iii) combined conformer selection/induced fit model. Physics-based molecular mechanics allowed the development of the force fields, which enabled the assessment of relative binding strength [155]. Alternative knowledge-based approaches were evolved to derive a statistical potential for interactions from a large database of protein-ligand complexes [156]. Affinity evaluation, statistical methods and combined procedures represent frequent alternatives of evaluation in recent protein docking approaches (Table 3). In addition to usual full 3D models simulating molecular complexes, contact docking was also proposed. This docking is based on contact map representations of molecules (cf. chapter 3). According to the authors contact docking appears feasible, and is able to complement other computational methods for the prediction of protein-protein interactions [157].

The number of molecules whose interactions can be scanned by current protein docking depends usually on the complexity of evaluated interactions and manner of simulation. For instance, interaction of 142 drugs inhibiting Poly-(ADP-ribose)-polymerase was tested in a large-scale virtual docking screening together with 300 000 added organic compounds with the aid of the program Lead Finder [158], whereas only 176 protein-protein interactions were analyzed by efficient program ZDOCK during extended simulation [159]. In addition to current output, some docking studies even verified the results of simulation using *in silico* experiments. For instance, aa involvement in enzyme interactions was proved with simulated mutation (program SYBYL 6.7 and Internal Coordinate Mechanics method [160, 161]) whereas other authors compared separate simulated effects of inhibitors and substrates (program BioDock [162]).

In addition to current molecular docking simulations (Table 3), certain docking studies analyzed modulation effects of additional molecules on the evaluated interaction [163–165]. Homology modeling followed by extensive molecular dynamics simulation was used to identify nonpeptidic small organic compounds (among 150 000 compounds) that bind to a human leukocyte antigen HLA-DR1301, and block the presentation of myelin basic protein peptide (aa positions 152–165) to T-cells. This peptide represents one of the epitopes critical for multiple sclerosis. *In silico* selection resulted in a set of 106 small molecules, two lead compounds were confirmed to specifically block IL-2 secretion by DR1301-restricted T-cells in a dose-dependent and reversible manner [163].

Pockets opening to protein surface represents potential sites for ligand binding or protein-protein interactions that were indeed identified in some cases [166–168]. Consequently, an algorithm BDOCK facilitating pocket-based prediction of protein binding site was proposed to improve protein docking [169], and moreover new algorithms predicting pockets are still proposed [170]. Many pockets were only transiently present on simulated surfaces of investigated proteins (e.g., Bcl-xl, interleukin 2, MDM2), and were all

opened only 2.5 picosecond when using model intervals corresponding to ten nanosecond range [171].

Some docking studies comprised phylogenic aspects of protein interactions. It was observed that except for antibody-antigen complexes [172], the surface density of conserved residue positions at the interface regions of interacting protein surfaces is high. The corresponding combination of the residue conservation information with a widely used shape complementarity algorithm improved the ability of protein docking to predict the native structure of protein-protein complexes. Efficient comparative docking consists in selection based on conservation in terms of chain positions and primary structure in the first step, and the following high-throughput structure-based docking evaluation [173] (see also Table 3). Recent combined strategy including multiple sequence alignment and fold recognition analysis has been proposed to perform more precise docking, and predict also protein function [174].

Bioinspired algorithms were also applied to molecular docking simulations including neural networks, evolutionary computing and swarm intelligence [175]. Though neural networks participated in some older evaluations working in conventional programs of molecular docking [176–179], their independent scoring functions for docking have appeared lately [180, 181]. An extreme neural network approach required only sequence input to perform a docking-like procedure [182]. Predictions followed from a trained model that simulated binding or nonbinding stages.

5. Protein Interaction Networks

Protein interaction networks (PINs) are usually represented by graphs. Nodes of these graphs denote interacting molecules, whereas each edge linking two nodes indicates the corresponding interaction. The prevailing part of PIN is usually constituted by protein-protein interaction network (P-PIN). In addition to P-PIN, we can observe a record of protein interactions with (i) nonpeptidic hormones or mediators and drugs, (ii) processed, targeted, and functional complexes forming RNA, and (iii) recombining, hypermutating, repairing, replicating, twisting/untwisting, and transcribed DNA. These nonpeptidic compounds represent in fact inputs, outputs, interaction-stabilizing agents, or relay batons of P-PIN. In cases of DNA, and RNA, some papers appeared dealing with special protein-RNA, and protein-DNA interaction networks [205–208]. Protein-DNA networks moreover combined both the protein-centric and the DNA-centric points of view [205].

Like in other related networks, clusters are recognized in P-PIN. Two main types of P-PIN-related clusters can be distinguished, that is, protein complexes and functional modules [209, 210]. Protein complexes are groups of proteins that interact with each other at the same time and place, forming a single multimolecular machine (e.g., metabolic multienzyme complexes). Functional modules consist of proteins that participate in a particular cellular process while binding to each other at a different time and place (e.g., multiple signaling cascades are functional modules

Table 3: Examples of protein docking.

Approach/combination	Investigated molecules	Predicted/investigated partners	Evaluation tools and results	Employed internet addresses	References
Protein-ligand					
SwissDock	protein	small molecule (ligand)	FullFitness + RMSD	http://www.swissdock.ch	[183]
Glide (version 5.6)	P-glycoprotein	24 P-gp binders + 102 endogenous molecules	Glide XP score + MM-GB/SA rescoring function	http://pubchem.ncbi.nlm.nih.gov/	[184]
VSDocker (AutoGrid4, AutoDock4, etc.)	protein	small molecule (example with 86 775 ligands)	ΔG_{bind}	http://www.bio.nnov.ru/projects/vsdocker2/	[185]
Opal (Autodock tools)	protein	small molecule (ligand)	ΔG_{bind}	http://kryptonite.ucsd.edu/opal2/dashboard?command=serviceList	[186]
FlexX module of SYBYL 7.0 and FMO	neuraminidase (from N1 subtype of influenza virus)	carboxyhexenyl derivatives or analogues	FlexX energy scores	http://www.rcsb.org/pdb	[187]
TarFisDock	protein	drugs	interaction energy	http://grid.ccnu.edu.cn/ http://www.dddc.ac.cn/tarfisdock/	[188]
KinDOCK	ATP binding sites of PK	focused ligand library	theoretical affinity	http://abcis.cbs.cnrs.fr/kindock	[189]
AutoDock + Gold	Cdc25 dual specificity phosphatases	library of sulfonylated aminothiazoles as inhibitors	ΔG_{bind} + overall Gold fitness function	none	[190]
Molsoft module + ACD database	thyroid hormone receptor	250 000 compounds	EQS score	none	[191]
Protein-protein					
ZDOCK	protein	flexible/shape complementary protein	combined evaluation using Fast Fourier Transform	http://zlab.umassmed.edu/zdock/ http://www.fftw.org	[192] [159, 193]
KBDOCK	protein	protein with hetero-binding site	combined knowledge-based approach	http://kbdock.loria.fr/	[194]
RosettaDock	protein	protein	CAPRI CS	http://rosettadock.graylab.jhu.edu	[195]
ClusPro	protein	protein	combined evaluation	http://cluspro.bu.edu/login.php	[196, 197]
Protein-nucleic acid					
DARS-RNP QUASI-RNP	protein	RNA	knowledge-based potentials for scoring protein-RNA models	http://iimcb.genesilico.pl/RNP/	[198]
protein-DNA docking approaches	transcription factor/protein	DNA	RMSD + knowledge based energy	http://pqs.ebi.ac.uk/ http://sparks.informatics.iupui.edu/czhang/ddna2.html	[199] [200]
Antigen receptors					
DS-QM based peptide binding prediction	HLA-DP2	HLA-DP2 interacting potential TCR ligands	CS following from two different DS-QM	none	[201]
SnugDock	antibody	antigen	paratope based structural optimization + CLEP	http://www.rosettacommons.org/	[202]
pDOCK	MHC I and II	potential TCR ligands	RMSD	http://biolinfo.org/mpid-t2	[203]
MHCsim	MHC I and II	potential TCR ligands	molecular dynamic simulation	http://igrid-ext.cryst.bbk.ac.uk/MHCsim/	[204]

ACD: available chemicals directory (database); CAPRI: critical assessment of predicted interactions; CLEP: combined lowest energy prediction; CS: combined scoring; DS-QM: docking score-based quantitative matrices; EQS score: score combining grid energy, electrostatic and entropy terms; FMO: fragment molecular orbitals; HLA-DP: a subset of human MHC II; ΔG_{bind}: free energy of binding; PK: protein kinases; MM-GB/SA molecular mechanics scoring function with generalized Born implicit solvent effects; P-gp: P-glycoprotein; RMSD: root mean square displacement statistics; TCR: T-cell receptor.

sometimes including also protein complexes). An important type of functional modules, that is, responsive functional modules (RFMs), includes protein interactions activated under specific conditions. These RFMs appear to be interesting with respect to prediction of potential biomarkers [211].

Much attention has been paid to the identification of small conserved subgraphs, particularly those occurring significantly often within the biological networks. These conserved subgraphs were referred to as network motifs or simply as motifs (similarly to primary structure motifs), and were considered as basic building blocks of complex networks including P-PIN [212, 213]. Ancestral pathways or functional modules including conserved interactions were derived when comparing P-PIN of phylogenetically distant species such as human, *Saccharomyces cerevisiae, Drosophila melanogaster* and *Caenorhabditis elegans*, and various bacterial species [214, 215]. Proteins necessary for proteasome function, transcription, RNA processing and translation were frequent in conserved subgraphs of compared *Eukaryota*, whereas DNA repair proteins prevailed in conserved prokaryotic subgraphs. Incorporation of literature-curated and evolutionarily conserved interactions also allowed to develop an interaction network for 54 proteins providing a tool for understanding Purkinje cell degeneration [216]. This result made it possible to find experimentally 770 mostly novel protein-protein interactions using a stringent yeast two-hybrid screen.

Recent integrative approaches combined at least transcriptional data with P-PIN analyses, and at most five layers including phenotype association with single-nucleotide polymorphism, disease-tissue, drug-tissue and drug-gene relationships, in addition to the topical P-PIN record [217] (see also Table 4). Two-layer comparison of disease-related mRNA expression, and human P-PIN was for instance used together with hierarchical clustering of networks to elucidate human disease similarities. The results led to a hypothesis considering common usage of some drugs in the diseases, which exhibited close relationship with respect to given clustering [218]. More extended analysis of gene expression overlays with protein-protein interaction, transcriptional-regulatory and signaling networks identified distinct driver-networks for each of the three common clinical breast cancer subtypes, that is, oestrogen receptor positive, human epidermal growth factor receptor 2 positive and triple receptor-negative breast cancers [219]. Integrative online accessible tools reevaluating P-PIN relationships were developed to predict candidate genes critical for the occurrence of different diseases (cf. network-based disease gene prioritization; [220, 221]).

The majority of proteins reciprocally interact via one or two interactions. This number substantially increases to tens, hundreds and more, when considering multi-interactive proteins denoted in PIN as hubs [222]. In agreement with the sense of the term, hubs are the principal agents in the interaction network and affect its function and stability [223]. Hubs were enriched in kinase and adaptor proteins including those interacting with frequently disordered phosphorylated protein regions [224, 225]. Similarly, many pathogenetically important proteins belonged

to hubs, for example, the widely known tumor suppressor p53, alpha-synuclein involved in Parkinson disease and small multifunctional core protein necessary for orchestration of viral progeny in *Flaviviridae* [226, 227]. In principle, two types of hubs were distinguished, that is, static "party hubs" and dynamic "date hubs." Party hubs are markedly more phylogenetically conserved, and their expression is highly correlated with their interacting partners in contrast to less conserved date hubs more frequently containing disordered regions and participating in cell signaling [228–230]. Hubs with two or more domains are more likely to connect distinct functional modules than single-domain hubs [224]. In addition to studies of biochemically interesting hubs, docking procedures (mediated by program ClusPro [231]) were recently employed together with comparative modeling to construct P-PIN [232], whereas domain-domain, domain-motif and motif-motif interactions were distinguished in some other P-PIN [233].

Extended P-PIN research yielded many tools allowing P-PIN-based prediction of protein complex formation [234–237], identification of hubs and multifunctional proteins [238, 239], identification of functional modules [240–242], network-based disease gene prioritization [217, 221, 243], as well as network building [244, 245], cross-species querying [246] or comparison [234, 237]. Some of these approaches as well as several examples of P-PIN-related databases are described in Table 4. The instrumental progress in PIN research also led to an increased number of the corresponding methodological reviews, for example, [247–250].

6. Accuracy as an Important Parameter

A detailed evaluation of accuracy of all the above approaches would require a separate review. The obstacles consist first of all in presence of complementary information in articles inaccessible for biologists and on web pages. Such state complicates mining of accuracy data first of all in the case, when server-related papers represent only the final step of author's efforts. Accuracy evaluation or even the corresponding references are also rarely present in the papers concerning new online databases. In addition, some authors also use other criterions related to accuracy to evaluate the corresponding performance value, whereas only some of such evaluations are widely known as valid accuracy substitution, for example, AUC (area under the ROC curve) discussed below.

In spite of the obstacles, we can mention several examples of high accuracy concerning the above approaches. Excellent accuracy values higher then 0.90 were found in the cases of Conserved Domain Database (or its RPS BLAST server; [257]; Table 1) and older 3D-BLAST variant under a broad range of conditions (Table 1), whereas similarly interesting AUC values (higher than 0.90) were mostly found also in the cases of recent 3D BLAST variant (Table 1), and the selected strategy of SVM- and machine-learning-based epitope prediction known as CBTOBE (chapter 3; [122]). The same accuracy levels were rarely achieved in certain subsets when employing ADAN, KinasePhos 2.0 (Table 1), ElliPro,

TABLE 4: Online accessible tools for PPI network assembly, reevaluation, and comparison.

	Purpose	Input	Evaluation results	Network content	Internet accessibility	References
Program tools						
PINTA	Resource for the prioritization of disease related candidate genes based on the differential expression of their neighborhood in a genome-wide protein-protein interaction network.	Files with disease-specific expression data	Internal scores, P values, Bonferroni–Holm P-values	menu: human, mouse, rat, worm (C. elegans) and yeast P-PIN	http://www.esat.kuleuven.be/pinta/	[221]
PCFamily	Finds homologous structure complexes of the query using BLASTP to search the structural template database	QS or QS set in FASTA format	Z-value agreement ratio	941 protein complexes	http://pcfamily.life.nctu.edu.tw	[236]
BisoGenet	Build and visualize biological networks in a fast and user-friendly manner including P-PIN	List of identifiers	Node degrees, cluster coefficients	5365 human genes	http://bio.cigb.edu.cu/bisogenet-cytoscape/	[245]
DynaMod	Identifies significant functional modules reflecting the change of modularity and differential expressions that are correlated with gene expression profiles under different conditions	Genome-wide expression profile	Z-score, Bonferroni or FDR P-values	In frame of GSEA	http://biocondor6.kaist.ac.kr/dynamod/	[241]
TORQUE	Cross species querying allows users to run topology-free queries on predefined or user-provided target networks resulting in subnetwork of the target network most similar to query subset	Seq of compared sets, query list, target P-PIN	Thickness of the edge	5430 proteins, 39 936 interactions	http://www.cs.tau.ac.il/~bnet/torque.html	[246]
NetworkBLAST	Identifies protein complexes within and across species, forming html page with links to schemes of interactions within complexes and additional graph data files	Files with *PPI, BL-PE *PPI	likelihood based density score	species related P-PIN *two species *single species	http://www.cs.tau.ac.il/~roded/networkblast.htm	[234]
Databases						
STRING	A database of functional interaction networks of proteins, globally integrated and scored	Search based on gene annotations	probabilistic confidence score	>1100 completely sequenced organisms	http://string-db.org	[251]
ChemProt	Disease chemical biology database including 7×10^5 chemicals and two millions chemical-protein interactions; indicates a possible formation of disease-associated protein complexes	Compound and protein identifiers	compilation of multiple chemical-protein annotations	30 578 proteins 2227 disease-related proteins, 428 429 PPI	http://www.cbs.dtu.dk/services/ChemProt/	[252]
NetAge	Database and network analysis tools for biogerontological research (integrity and functionality of P-PIN is under a tight epigenetic control)	Tools for searching and browsing	experimental evidence about interactions	miRNA-regulated P-PIN in age-related diseases	http://www.netage-project.org	[253]
BioGRID	All interactions in BioGRID are available through the Osprey visualization system, which can be used to query network organization in a user-defined fashion	Intuitive graphical interface	experimental evidence about interactions	Over 198 000 genetic + protein interactions from six species	http://www.thebiogrid.org	[254]

TABLE 4: Continued.

	Purpose	Input	Evaluation results	Network content	Internet accessibility	References
EHCO	Encyclopedia of Hepatocellular Carcinoma genes Online collect, organize and compare unsorted HCC-related studies	CMS	NL processing and softbots	97 proteins, 47 highly interactive, 18 hubs	http://ehco.iis.sinica.edu.tw	[255]
IntNetDB v1.0	A database providing automatic prediction and visualization of PPI network among genes of interest; analysis includes domain-domain interactions, known gene contexts, crossvalidation, and so forth	list of query gene identifiers	likelihood ratio following from Bayesian analysis	concerns 27 species contains GSP and GSN from HPRD	http://hanlab.genetics.ac.cn/IntNetDB.htm	[256]

*Independent alternatives; BLPE: BLASTP *E* values between pairs of proteins from each of the compared species; CMS: content management system; GSEA: gene set enrichment analyses; GSP, GSN: gold standard positive and a gold standard negative dataset of HPRD, respectively; HCC: hepatocellular carcinoma; HPRD: human protein reference database; NL: natural language; PE: probabilistic evaluation; PPI: protein-protein interactions; QS: query sequences; seq: sequences.

RANKPEP (Table 2), SVM and machine-learning-based epitope prediction improved by evolutionary information BEOracle (chapter 3; [124]) and docking tool HADDOCK (chapter 4; [178]). Similarly, restricted subset-related AUC values of 0.90 and 0.93 indicated a good predicting ability when using docking-like Partner-aware prediction (chapter 4; [182]), and combined evaluation based on Glide X and molecular mechanics (Table 3) to simulate interactions of 25 antibodies and P-glycoprotein, respectively. Frequent subset accuracies of higher values than the gold standard value of 0.80 were observed during runs of programs Phos3D, KinasePhos 2.0 (Table 1), ElliPro (Table 2) and CBTOBE mentioned above (chapter 3; [122]), and two docking tools Glide XP and Gold (Table 2; [258]). Comparable AUC values then concerned RANKPEP (Table 2), BEOracle mentioned above (chapter 3; [124]) and network-based disease gene prioritization called DADA (chapter 5; [236]). The low occurrence of P-PIN in our overview of accuracy evaluation followed from the fact that the performance of P-PIN was alternatively evaluated with the aid of precision or recall. An interesting approach—Lead Finder—was developed to improve the accuracy of protein-ligand docking, binding energy estimation and virtual screening. High enrichment factors were obtained for almost all of the targets and seven different docking programs resulting in an excellent average AUC value of 0.92 [258].

7. Tendencies, Improvements, and Interplays

In summary, virtual interactomics represents a highly sophisticated, and relatively autonomous subarea of both interactomics and bioinformatics. Its development can be characterized by several independent and integrative tendencies. The independent tendencies include (i) more rapid and precise molecular modeling of protein complexes (including quantum mechanic methods and rapid pharmacological screening based on docking), (ii) more complex evaluation of structural similarities (e.g., large alphabets and combined prediction of functional sites), (iii) extensive assembly of large-scale PIN and assembly of PIN or PIN subgraphs with dynamical properties, and (iv) restriction of conserved PIN subgraphs and clinically interesting responsive functional modules. The integrative tendencies then comprise (i) establishment of structurally functionally, and purposefully oriented databases and associated data mining servers, (ii) questions concerning phylogenic relationships between structural differences in PIN subgraphs, and the differences in protein structures of subgraph forming molecules, (iii) PIN prediction and revision/reevaluation of empirically proposed PIN using comparative molecular modeling, docking and bio-inspired algorithms, (iv) comparative differential multilayer network approaches including clinical treatment cases, and (v) attempts to construct global networks within a single cell, and hopefully, within the system of different reciprocally communicating cell types (e.g., immune system). Since interactomics simplifies reactomic simulations, it can also markedly and rapidly complete reactomic results, for example, in prediction of enzyme-inhibitor effects and

specificity of protein kinase interactions mentioned above. In the end, we also dealt here with accuracy, which represents an important parameter of tool choice together with reasonable tool accessibility, and possibilities to test several alternatives before prediction, simulation and network analysis or to use two or several independent methods. In conclusion, many interesting and important interplays occur inside the virtual interactomic area as well as in its boundary lines and performance evaluation.

Abbreviations

3D: Three-dimensional
aa: Amino acid residues in protein sequence
PIN: Protein interaction network(s)
P-PIN: Protein-protein interaction network(s).

References

[1] A. Espejo, J. Côté, A. Bednarek, S. Richard, and M. T. Bedford, "A protein-domain microarray identifies novel protein-protein interactions," *Biochemical Journal*, vol. 367, no. 3, pp. 697–702, 2002.

[2] S. H. Lo, "Reverse interactomics: from peptides to proteins and to functions," *ACS Chemical Biology*, vol. 2, no. 2, pp. 93–95, 2007.

[3] B. Suter, S. Kittanakom, and I. Stagljar, "Two-hybrid technologies in proteomics research," *Current Opinion in Biotechnology*, vol. 19, no. 4, pp. 316–323, 2008.

[4] D. J. Briant, J. M. Murphy, G. C. Leung, and F. Sicheri, "Rapid identification of linear protein domain binding motifs using peptide SPOT arrays," *Methods in Molecular Biology*, vol. 570, pp. 175–185, 2009.

[5] S. Hui and G. D. Bader, "Proteome scanning to predict PDZ domain interactions using support vector machines," *BMC Bioinformatics*, vol. 11, article 507, 2010.

[6] A. Daemen, M. Signoretto, O. Gevaert, J. A. K. Suykens, and B. De Moor, "Improved microarray-based decision support with graph encoded interactome data," *PLoS ONE*, vol. 5, no. 4, Article ID e10225, 2010.

[7] J. Huang, B. Ru, and P. Dai, "Bioinformatics resources and tools for phage display," *Molecules*, vol. 16, no. 1, pp. 694–709, 2011.

[8] A. D. J. van Dijk, R. Kaptein, R. Boelens, and A. M. J. J. Bonvin, "Combining NMR relaxation with chemical shift perturbation data to drive protein-protein docking," *Journal of Biomolecular NMR*, vol. 34, no. 4, pp. 237–244, 2006.

[9] S. Klinge, R. Ñez-Ramírez, O. Llorca, and L. Pellegrini, "3D architecture of DNA Pol α reveals the functional core of multi-subunit replicative polymerases," *The EMBO Journal*, vol. 28, no. 13, pp. 1978–1987, 2009.

[10] J. L. Stark and R. Powers, "Application of NMR and molecular docking in structure-based drug discovery," *Topics in Current Chemistry*, vol. 326, pp. 1–34, 2012.

[11] J. Orts, S. Bartoschek, C. Griesinger, P. Monecke, and T. Carlomagno, "An NMR-based scoring function improves the accuracy of binding pose predictions by docking by two orders of magnitude," *Journal of Biomolecular NMR*, vol. 52, no. 1, pp. 23–30, 2012.

[12] D. J. Diller and R. Li, "Kinases, homology models, and high throughput docking," *Journal of Medicinal Chemistry*, vol. 46, no. 22, pp. 4638–4647, 2003.

[13] F. Förster and E. Villa, "Integration of cryo-EM with atomic and protein-protein interaction data," *Methods in Enzymology*, vol. 483, pp. 47–72, 2010.

[14] C. N. Cavasotto, "Homology models in docking and high-throughput docking," *Current Topics in Medicinal Chemistry*, vol. 11, no. 12, pp. 1528–1534, 2011.

[15] Y. Tsfadia, R. Friedman, J. Kadmon, A. Selzer, E. Nachliel, and M. Gutman, "Molecular dynamics simulations of palmitate entry into the hydrophobic pocket of the fatty acid binding protein," *FEBS Letters*, vol. 581, no. 6, pp. 1243–1247, 2007.

[16] M. Gochin, G. Zhou, and A. H. Phillips, "Paramagnetic relaxation assisted docking of a small indole compound in the HIV-1 gp41 hydrophobic pocket," *ACS Chemical Biology*, vol. 6, no. 3, pp. 267–274, 2011.

[17] H. Blankenburg, F. Ramírez, J. Büch, and M. Albrecht, "DAS-MIweb: online integration, analysis and assessment of distributed protein interaction data," *Nucleic Acids Research*, vol. 37, no. 2, pp. W122–W128, 2009.

[18] M. J. Cowley, M. Pinese, K. S. Kassahn et al., "PINA v2.0: mining interactome modules," *Nucleic Acids Research*, vol. 40, pp. D862–D865, 2011.

[19] M. Sickmeier, J. A. Hamilton, T. LeGall et al., "DisProt: the database of disordered proteins," *Nucleic Acids Research*, vol. 35, no. 1, pp. D786–D793, 2007.

[20] B. Meszaros, I. Simon, Z. Dosztanyi, et al., "Prediction of protein binding regions in disordered proteins," *PLoS Computational Biology*, vol. 5, no. 5, Article ID e1000376, 2009.

[21] B. Mészáros, I. Simon, and Z. Dosztányi, "The expanding view of protein-protein interactions: complexes involving intrinsically disordered proteins," *Physical Biology*, vol. 8, no. 3, Article ID 035003, 2011.

[22] A. Biagini and A. Puigserver, "Sequence analysis of the aminoacylase-1 family. A new proposed signature for metalloexopeptidases," *Comparative Biochemistry and Physiology B*, vol. 128, no. 3, pp. 469–481, 2001.

[23] J. Kubrycht, K. Sigler, M. Růžička, P. Souček, J. Borecký, and P. Ježek, "Ancient phylogenetic beginnings of immunoglobulin hypermutation," *Journal of Molecular Evolution*, vol. 63, no. 5, pp. 691–706, 2006.

[24] D. Przybylski and B. Rost, "Consensus sequences improve PSI-BLAST through mimicking profile-profile alignments," *Nucleic Acids Research*, vol. 35, no. 7, pp. 2238–2246, 2007.

[25] D. Przybylski and B. Rost, "Powerful fusion: PSI-BLAST and consensus sequences," *Bioinformatics*, vol. 24, no. 18, pp. 1987–1993, 2008.

[26] K. O. Wrzeszczynski and B. Rost, "Cell cycle kinases predicted from conserved biophysical properties," *Proteins*, vol. 74, no. 3, pp. 655–668, 2009.

[27] K. Beck, I. Hunter, and J. Engel, "Structure and function of laminin: anatomy of a multidomain glycoprotein," *The FASEB Journal*, vol. 4, no. 2, pp. 148–160, 1990.

[28] X.-N. Dong, Y. Xiao, M. P. Dierich, and Y. H. Chen, "N- and C-domains of HIV-1 gp41: mutation, structure and functions," *Immunology Letters*, vol. 75, no. 3, pp. 215–220, 2001.

[29] P. R. Sibbald, H. Sommerfeldt, and P. Argos, "Automated protein sequence pattern handling and PROSITE searching," *Computer Applications in the Biosciences*, vol. 7, no. 4, pp. 535–536, 1991.

[30] A. Gattiker, E. Gasteiger, and A. Bairoch, "ScanProsite: a reference implementation of a PROSITE scanning tool," *Appl Bioinformatics*, vol. 1, no. 2, pp. 107–108, 2002.

[31] C. J. A. Sigrist, E. De Castro, P. S. Langendijk-Genevaux, V. Le Saux, A. Bairoch, and N. Hulo, "ProRule: a new database containing functional and structural information on PROSITE profiles," *Bioinformatics*, vol. 21, no. 21, pp. 4060–4066, 2005.

[32] C. J. A. Sigrist, L. Cerutti, E. De Castro et al., "PROSITE, a protein domain database for functional characterization and annotation," *Nucleic Acids Research*, vol. 38, no. 1, Article ID gkp885, pp. D161–D166, 2009.

[33] R. L. Dunbrack Jr., "Comparative modeling of CASP3 targets using PSI-BLAST and SCWRL," *Proteins*, supplement 3, pp. 81–87, 1999.

[34] A. Elofsson, "A study on protein sequence alignment quality," *Proteins*, vol. 46, no. 3, pp. 330–339, 2002.

[35] H. Zhou and Y. Zhou, "Single-body residue-level knowledge-based energy score combined with sequence-profile and secondary structure information for fold recognition," *Proteins*, vol. 55, no. 4, pp. 1005–1013, 2004.

[36] L. Jaroszewski, L. Rychlewski, Z. Li, W. Li, and A. Godzik, "FFAS03: a server for profile-profile sequence alignments," *Nucleic Acids Research*, vol. 33, no. 2, pp. W284–W288, 2005.

[37] L. Jaroszewski, Z. Li, X. H. Cai, C. Weber, and A. Godzik, "FFAS server: novel features and applications," *Nucleic Acids Research*, vol. 39, no. 2, pp. W38–W44, 2011.

[38] E. Dong, J. Smith, S. Heinze, N. Alexander, and J. Meiler, "BCL::Align-Sequence alignment and fold recognition with a custom scoring function online," *Gene*, vol. 422, no. 1-2, pp. 41–46, 2008.

[39] K. Fallen, S. Banerjee, J. Sheehan et al., "The Kir channel immunoglobulin domain is essential for Kir1.1 (ROMK) thermodynamic stability, trafficking and gating," *Channels*, vol. 3, no. 1, pp. 57–68, 2009.

[40] B. Vishnepolsky and M. Pirtskhalava, "CONTSOR—a new knowledge-based fold recognition potential, based on side chain orientation and contacts between residue terminal groups," *Protein Science*, vol. 21, no. 1, pp. 134–141, 2012.

[41] J.-M. Yang and C. H. Tung, "Protein structure database search and evolutionary classification," *Nucleic Acids Research*, vol. 34, no. 13, pp. 3646–3659, 2006.

[42] L. Mavridis, A. W. Ghoorah, V. Venkatraman, and D. W. Ritchie, "Representing and comparing protein folds and fold families using three-dimensional shape-density representations," *Proteins*, vol. 80, no. 2, pp. 530–545, 2012.

[43] P. Di Lena, P. Fariselli, L. Margara, M. Vassura, and R. Casadio, "Fast overlapping of protein contact maps by alignment of eigenvectors," *Bioinformatics*, vol. 26, no. 18, pp. 2250–2258, 2010.

[44] M. A. Rodionov and M. S. Johnson, "Residue-residue contact substitution probabilities derived from aligned three-dimensional structures and the identification of common folds," *Protein Science*, vol. 3, no. 12, pp. 2366–2377, 1994.

[45] A. Godzik, J. Skolnick, and A. Kolinski, "Regularities interaction patterns of globular proteins," *Protein Engineering*, vol. 6, no. 8, pp. 801–810, 1993.

[46] C. Vehlow, H. Stehr, M. Winkelmann et al., "CMView: interactive contact map visualization and analysis," *Bioinformatics*, vol. 27, no. 11, Article ID btr163, pp. 1573–1574, 2011.

[47] J. A. Barker and J. M. Thornton, "An algorithm for constraint-based structural template matching: application to 3D templates with statistical analysis," *Bioinformatics*, vol. 19, no. 13, pp. 1644–1649, 2003.

[48] R. M. Ward, E. Venner, B. Daines et al., "Evolutionary trace annotation server: automated enzyme function prediction in

protein structures using 3D templates," *Bioinformatics*, vol. 25, no. 11, pp. 1426–1427, 2009.

[49] J. C. Nebel, "Generation of 3D templates of active sites of proteins with rigid prosthetic groups," *Bioinformatics*, vol. 22, no. 10, pp. 1183–1189, 2006.

[50] P. P. Kuksa and V. Pavlovic, "Efficient motif finding algorithms for large-alphabet inputs," *BMC Bioinformatics*, vol. 11, supplement 8, article S1, 2010.

[51] O. Dror, H. Benyamini, R. Nussinov, and H. Wolfson, "MASS: multiple structural alignment by secondary structures," *Bioinformatics*, vol. 19, no. 1, pp. i95–i104, 2003.

[52] O. Dror, H. Benyamini, R. Nussinov, and H. J. Wolfson, "Multiple structural alignment by secondary structures: algorithm and applications," *Protein Science*, vol. 12, no. 11, pp. 2492–2507, 2003.

[53] A. Williams, D. R. Gilbert, and D. R. Westhead, "Multiple structural alignment for distantly related all β structures using TOPS pattern discovery and simulated annealing," *Protein Engineering*, vol. 16, no. 12, pp. 913–923, 2003.

[54] A. S. Konagurthu, J. C. Whisstock, P. J. Stuckey, and A. M. Lesk, "MUSTANG: a multiple structural alignment algorithm," *Proteins*, vol. 64, no. 3, pp. 559–574, 2006.

[55] C. Micheletti and H. Orland, "MISTRAL: a tool for energy-based multiple structural alignment of proteins," *Bioinformatics*, vol. 25, no. 20, pp. 2663–2669, 2009.

[56] A. S. Konagurthu, C. F. Reboul, J. W. Schmidberger et al., "MUSTANG-MR structural sieving server: applications in protein structural analysis and crystallography," *PLoS ONE*, vol. 5, no. 4, Article ID e10048, 2010.

[57] W. Y. Siu, N. Mamoulis, S. M. Yiu, and H. L. Chan, "A data-mining approach for multiple structural alignment of proteins," *Bioinformation*, vol. 4, no. 8, pp. 366–370, 2010.

[58] H. Sun, A. Sacan, H. Ferhatosmanoglu, and Y. Wang, "Smolign: a spatial motifs based protein multiple structural alignment method," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 9, no. 1, 14 pages, 2012.

[59] M. L. Miller, L. J. Jensen, F. Diella et al., "Linear motif atlas for phosphorylation-dependent signaling," *Science Signaling*, vol. 1, no. 35, p. ra2, 2008.

[60] D. Schwartz and G. M. Church, "Collection and motif-based prediction of phosphorylation sites in human viruses," *Science Signaling*, vol. 3, no. 137, p. rs2, 2010.

[61] D. M. Standley, R. Yamashita, A. R. Kinjo, H. Toh, and H. Nakamura, "SeSAW: balancing sequence and structural information in protein functional mapping," *Bioinformatics*, vol. 26, no. 9, pp. 1258–1259, 2010.

[62] Y. H. Wong, T. Y. Lee, H. K. Liang et al., "KinasePhos 2.0: a web server for identifying protein kinase-specific phosphorylation sites based on sequences and coupling patterns," *Nucleic acids research*, vol. 35, pp. W588–W594, 2007.

[63] P. Durek, C. Schudoma, W. Weckwerth, J. Selbig, and D. Walther, "Detection and characterization of 3D-signature phosphorylation site motifs and their contribution towards improved phosphorylation site prediction in proteins," *BMC Bioinformatics*, vol. 10, article 117, 2009.

[64] J. A. Encinar, G. Fernandez-Ballester, I. E. Sánchez et al., "ADAN: a database for prediction of protein-protein interaction of modular domains mediated by linear motifs," *Bioinformatics*, vol. 25, no. 18, pp. 2418–2424, 2009.

[65] A. Marchler-Bauer, A. R. Panchenko, B. A. Shoemarker, P. A. Thiessen, L. Y. Geer, and S. H. Bryant, "CDD: a database of conserved domain alignments with links to domain three-dimensional structure," *Nucleic Acids Research*, vol. 30, no. 1, pp. 281–283, 2002.

[66] A. Marchler-Bauer, J. B. Anderson, F. Chitsaz et al., "CDD: specific functional annotation with the Conserved Domain Database," *Nucleic Acids Research*, vol. 37, no. 1, pp. D205–D210, 2009.

[67] A. Marchler-Bauer, S. Lu, J. B. Anderson et al., "CDD: a Conserved Domain Database for the functional annotation of proteins," *Nucleic Acids Research*, vol. 39, no. 1, pp. D225–D229, 2011.

[68] P. J. Kundrotas, Z. Zhu, and I. A. Vakser, "GWIDD: genome-wide protein docking database," *Nucleic Acids Research*, vol. 38, no. 1, pp. D513–D517, 2009.

[69] G. Pugalenthi, P. N. Suganthan, R. Sowdhamini, and S. Chakrabarti, "MegaMotifBase: a database of structural motifs in protein families and superfamilies," *Nucleic Acids Research*, vol. 36, no. 1, pp. D218–D221, 2008.

[70] P. May, A. Kreuchwig, T. Steinke, and I. Koch, "PTGL: a database for secondary structure-based protein topologies," *Nucleic Acids Research*, vol. 38, no. 1, pp. D326–D330, 2010.

[71] A. Shulman-Peleg, R. Nussinov, and H. J. Wolfson, "RsiteDB: a database of protein binding pockets that interact with RNA nucleotide bases," *Nucleic Acids Research*, vol. 37, no. 1, pp. D369–D373, 2009.

[72] H. M. Geysen, S. J. Rodda, and T. J. Mason, "A priori delineation of a peptide which mimics a discontinuous antigenic determinant," *Molecular Immunology*, vol. 23, no. 7, pp. 709–715, 1986.

[73] J. M. Penninger and K. Bachmaier, "Review of microbial infections and the immune response to cardiac antigens," *Journal of Infectious Diseases*, vol. 181, no. 6, pp. S498–S504, 2000.

[74] N. Fernandez-Fuentes, B. Oliva, and A. Fiser, "A supersecondary structure library and search algorithm for modeling loops in protein structures," *Nucleic Acids Research*, vol. 34, no. 7, pp. 2085–2097, 2006.

[75] J. E. Libbey, L. L. McCoy, and R. S. Fujinami, "Molecular mimicry in multiple sclerosis," *International Review of Neurobiology*, vol. 79, pp. 127–147, 2007.

[76] G. P. Smith, "Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface," *Science*, vol. 228, no. 4705, pp. 1315–1317, 1985.

[77] J. J. Devlin, L. C. Panganiban, and P. E. Devlin, "Random peptide libraries: a source of specific protein binding molecules," *Science*, vol. 249, no. 4967, pp. 404–406, 1990.

[78] J. K. Scott and G. P. Smith, "Searching for peptide ligands with an epitope library," *Science*, vol. 249, no. 4967, pp. 386–390, 1990.

[79] C. L. Jellis, T. J. Cradick, P. Rennert et al., "Defining critical residues in the epitope for a HIV-neutralizing monoclonal antibody using phage display and peptide array technologies," *Gene*, vol. 137, no. 1, pp. 63–68, 1993.

[80] D. J. Matthews and J. A. Wells, "Substrate phage: selection of protease substrates by monovalent phage display," *Science*, vol. 260, no. 5111, pp. 1113–1117, 1993.

[81] E. Koivunen, B. Wang, and E. Ruoslahti, "Isolation of a highly specific ligand for the $\alpha 5\beta 1$ integrin from a phage display library," *Journal of Cell Biology*, vol. 124, no. 3, pp. 373–380, 1994.

[82] R. Pasqualini, E. Koivunen, and E. Ruoslahti, "A peptide isolated from phage display libraries is a structural and functional mimic of an RGD-binding site on integrins," *Journal of Cell Biology*, vol. 130, no. 5, pp. 1189–1196, 1995.

[83] R. Schmitz, G. Baumann, and H. Gram, "Catalytic specificity of phosphotyrosine kinases Blk, Lyn, c-Src and Syk as assessed by phage display," *Journal of Molecular Biology*, vol. 260, no. 5, pp. 664–677, 1996.

[84] F. Fack, B. Hügle-Dörr, D. Song, I. Queitsch, G. Petersen, and E. K. F. Bautz, "Epitope mapping by phage display: random versus gene-fragment libraries," *Journal of Immunological Methods*, vol. 206, no. 1-2, pp. 43–52, 1997.

[85] M. Blüthner, C. Schäfer, C. Schneider, and F. A. Bautz, "Identification of major linear epitopes on the sp100 nuclear PBC autoantigen by the gene-fragment phage-display technology," *Autoimmunity*, vol. 29, no. 1, pp. 33–42, 1999.

[86] R. C. Ladner, A. K. Sato, J. Gorzelany, and M. de Souza, "Phage display-derived peptides as therapeutic alternatives to antibodies," *Drug Discovery Today*, vol. 9, no. 12, pp. 525–529, 2004.

[87] R. Brissette, J. K. A. Prendergast, and N. I. Goldstein, "Identification of cancer targets and therapeutics using phage display," *Current Opinion in Drug Discovery and Development*, vol. 9, no. 3, pp. 363–369, 2006.

[88] T. Hohm, P. Limbourg, and D. Hoffmann, "A multiobjective evolutionary method for the design of peptidic mimotopes," *Journal of Computational Biology*, vol. 13, no. 1, pp. 113–125, 2006.

[89] G. Castel, M. Chtéoui, B. Heyd, and N. Tordo, "Phage display of combinatorial peptide libraries: application to antiviral research," *Molecules*, vol. 16, no. 5, pp. 3499–3518, 2011.

[90] T. Sherev, K. H. Wiesmüller, and P. Walden, "Mimotopes of tumor-associated T-cell epitopes for cancer vaccines determined with combinatorial peptide libraries," *Applied Biochemistry and Biotechnology B*, vol. 25, no. 1, pp. 53–61, 2003.

[91] Y. W. Zhong, D. P. Xu, X. D. Li, J. Z. Dai, B. Xu, and L. Li, "Epitope screening of influenza A (H3N2) by using phage display library," *Zhonghua Shi Yan He Lin Chuang Bing Du Xue Za Zhi*, vol. 23, no. 4, pp. 272–274, 2009.

[92] E.-C. Sun, J. Zhao, T. Yang et al., "Identification of a conserved JEV serocomplex B-cell epitope by screening a phage-display peptide library with a mAb generated against West Nile virus capsid protein," *Virology Journal*, vol. 8, no. 100, pp. 1–10, 2011.

[93] D. A. Rothenfluh, H. Bermudez, C. P. O'Neil, and J. A. Hubbell, "Biofunctional polymer nanoparticles for intra-articular targeting and retention in cartilage," *Nature Materials*, vol. 7, no. 3, pp. 248–254, 2008.

[94] R. J. Giordano, J. K. Edwards, R. M. Tuder, W. Arap, and R. Pasqualini, "Combinatorial ligand-directed lung targeting," *Proceedings of the American Thoracic Society*, vol. 6, no. 5, pp. 411–415, 2009.

[95] R. J. Passarella, D. E. Spratt, A. E. van der Ende et al., "Targeted nanoparticles that deliver a sustained, specific release of paclitaxel to irradiated tumors," *Cancer Research*, vol. 70, no. 11, pp. 4550–4559, 2010.

[96] G. Y. Lee, J. H. Kim, G. T. Oh, B. H. Lee, I. C. Kwon, and I. S. Kim, "Molecular targeting of atherosclerotic plaques by a stabilin-2-specific peptide ligand," *Journal of Controlled Release*, vol. 155, no. 2, pp. 211–217, 2011.

[97] J. Li, L. Feng, L. Fan et al., "Targeting the brain with PEG-PLGA nanoparticles modified with phage-displayed peptides," *Biomaterials*, vol. 32, no. 21, pp. 4943–4950, 2011.

[98] V. P. Valuev, D. A. Afonnikov, M. P. Ponomarenko, L. Milanesi, and N. A. Kolchanov, "ASPD (Artificially Selected Proteins/Peptides Database): a database of proteins and

[99] peptides evolved in vitro," *Nucleic Acids Research*, vol. 30, no. 1, pp. 200–202, 2002.

[99] S. Mandava, L. Makowski, S. Devarapalli, J. Uzubell, and D. J. Rodi, "RELIC—a bioinformatics server for combinatorial peptide analysis and identification of protein-ligand interaction sites," *Proteomics*, vol. 4, no. 5, pp. 1439–1460, 2004.

[100] J. L. Casey, A. M. Coley, and M. Foley, "Phage display of peptides in ligand selection for use in affinity chromatography," *Methods in Molecular Biology*, vol. 421, pp. 111–124, 2007.

[101] B. M. Mumey, B. W. Bailey, B. Kirkpatrick, A. J. Jesaitis, T. Angel, and E. A. Dratz, "A new method for mapping discontinuous antibody epitopes to reveal structural features of proteins," *Journal of Computational Biology*, vol. 10, no. 3-4, pp. 555–567, 2003.

[102] B. Mumey, N. Ohler, T. Angel, A. Jesaitis, and E. Dratz, "Filtering epitope alignments to improve protein surface prediction," *Frontiers of High Performance Computing and Networking*, vol. 4331, pp. 648–657, 2006.

[103] V. Moreau, C. Granier, S. Villard, D. Laune, and F. Molina, "Discontinuous epitope prediction based on mimotope analysis," *Bioinformatics*, vol. 22, no. 9, pp. 1088–1095, 2006.

[104] B. Ru, J. Huang, P. Dai et al., "MimoDB: a new repository for mimotope data derived from phage display technology," *Molecules*, vol. 15, no. 11, pp. 8279–8288, 2010.

[105] J. Huang, B. Ru, P. Zhu et al., "MimoDB 2.0: a mimotope database and beyond," *Nucleic Acids Research*, vol. 40, pp. D271–D277, 2012.

[106] I. Mayrose, T. Shlomi, N. D. Rubinstein et al., "Epitope mapping using combinatorial phage-display libraries: a graph-based algorithm," *Nucleic Acids Research*, vol. 35, no. 1, pp. 69–78, 2007.

[107] Y. X. Huang, Y. L. Bao, S. Y. Guo, Y. Wang, C. G. Zhou, and Y. X. Li, "Pep-3D-Search: a method for B-cell epitope prediction based on mimotope analysis," *BMC Bioinformatics*, vol. 9, article 538, 2008.

[108] W. H. Chen, P. P. Sun, Y. Lu, W. W. Guo, Y. X. Huang, and Z. Q. Ma, "MimoPro: a more efficient Web-based tool for epitope prediction using phage display libraries," *BMC Bioinformatics*, vol. 12, article 199, 2011.

[109] G. F. Denisova, D. A. Denisov, and J. L. Bramson, "Applying bioinformatics for antibody epitope prediction using affinity-selected mimotopes—relevance for vaccine design," *Immunome Research*, vol. 6, supplement 2, article S6, 2010.

[110] C. J. Bryson, T. D. Jones, and M. P. Baker, "Prediction of immunogenicity of therapeutic proteins: validity of computational tools," *BioDrugs*, vol. 24, no. 1, pp. 1–8, 2010.

[111] P. Sun, W. Chen, Y. Huang, H. Wang, Z. Ma, and Y. Lv, "Epitope prediction based on random peptide library screening: benchmark dataset and prediction tools evaluation," *Molecules*, vol. 16, no. 6, pp. 4971–4993, 2011.

[112] U. Kulkarni-Kale, S. Bhosle, and A. S. Kolaskar, "CEP: a conformational epitope prediction server," *Nucleic Acids Research*, vol. 33, no. 2, pp. W168–W171, 2005.

[113] V. Moreau, C. Fleury, D. Piquer et al., "PEPOP: computational design of immunogenic peptides," *BMC Bioinformatics*, vol. 9, article 71, 2008.

[114] J. Ponomarenko, H.-H. Bui, W. Li et al., "ElliPro: a new structure-based tool for the prediction of antibody epitopes," *BMC Bioinformatics*, vol. 9, article 514, 2008.

[115] N. D. Rubinstein, I. Mayrose, E. Martz, and T. Pupko, "Epitopia: a web-server for predicting B-cell epitopes," *BMC Bioinformatics*, vol. 10, article 287, 2009.

[116] J. Van de Water, M. E. Gershwin, P. Leung, A. Ansari, and R. L. Coppel, "The autoepitope of the 74-kD mitochondrial

autoantigen of primary biliary cirrhosis corresponds to the functional site of dihydrolipoamide acetyltransferase," *Journal of Experimental Medicine*, vol. 167, no. 6, pp. 1791–1799, 1988.

[117] R. F. El-Kased, C. Koy, T. Deierling et al., "Mass spectrometric and peptide chip epitope mapping of rheumatoid arthritis autoantigen RA33," *European Journal of Mass Spectrometry*, vol. 15, no. 6, pp. 747–759, 2009.

[118] V. Burkart, R. K. Siegenthaler, E. Blasius et al., "High affinity binding of hydrophobic and autoantigenic regions of proinsulin to the 70 kDa chaperone DnaK," *BMC Biochemistry*, vol. 11, no. 1, article 44, 2010.

[119] S. Liang, D. Zheng, D. M. Standley, B. Yao, M. Zacharias, and C. Zhang, "EPSVR and EPMeta: prediction of antigenic epitopes using support vector regression and multiple server results," *BMC Bioinformatics*, vol. 11, article 381, 2010.

[120] L. J. K. Wee, D. Simarmata, Y. W. Kam, L. F. P. Ng, and J. C. Tong, "SVM-based prediction of linear B-cell epitopes using Bayes Feature Extraction," *BMC Genomics*, vol. 11, supplement 4, article S21, 2010.

[121] C.-H. Su, N. R. Pal, K.-L. Lin, and I.-F. Chung, "Identification of amino acid propensities that are strong determinants of linear B-cell epitope using neural networks," *PLoS ONE*, vol. 7, no. 2, Article ID e30617, 2012.

[122] H. R. Ansari and G. P. Raghava, "Identification of conformational B-cell Epitopes in an antigen from its primary sequence," *Immunome Research*, vol. 6, no. 1, article 6, 2010.

[123] M. J. Sweredoski and P. Baldi, "COBEpro: a novel system for predicting continuous B-cell epitopes," *Protein Engineering, Design and Selection*, vol. 22, no. 3, pp. 113–120, 2009.

[124] Y. Wang, W. Wu, N. N. Negre, K. P. White, C. Li, and P. K. Shah, "Determinants of antigenicity and specificity in immune response for protein sequences," *BMC Bioinformatics*, vol. 12, article 251, 2011.

[125] M. V. Larsen, C. Lundegaard, K. Lamberth et al., "An integrative approach to CTL epitope prediction: a combined algorithm integrating MHC class I binding, TAP transport efficiency, and proteasomal cleavage predictions," *European Journal of Immunology*, vol. 35, no. 8, pp. 2295–2303, 2005.

[126] M. V. Larsen, C. Lundegaard, K. Lamberth, S. Buus, O. Lund, and M. Nielsen, "Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction," *BMC Bioinformatics*, vol. 8, article 424, 2007.

[127] I. A. Doytchinova, P. Guan, and D. R. Flower, "EpiJen: a server for multistep T cell epitope prediction," *BMC Bioinformatics*, vol. 7, article 131, 2006.

[128] M. Feldhahn, P. Dönnes, P. Thiel, and O. Kohlbacher, "FRED—a framework for T-cell epitope detection," *Bioinformatics*, vol. 25, no. 20, pp. 2758–2759, 2009.

[129] C. K. Hattotuwagama, C. P. Toseland, P. Guan et al., "Toward prediction of class II mouse major histocompatibility complex peptide binding affinity: in silico bioinformatic evaluation using partial least squares, a robust multivariate statistical technique," *Journal of Chemical Information and Modeling*, vol. 46, no. 3, pp. 1491–1502, 2006.

[130] C. Lundegaard, O. Lund, and M. Nielsen, "Prediction of epitopes using neural network based methods," *Journal of Immunological Methods*, vol. 374, no. 1-2, pp. 26–34, 2011.

[131] C.-W. Tung, M. Ziehm, A. Kämper, O. Kohlbacher, and S.-Y. Ho, "POPISK: T-cell reactivity prediction using support vector machines and string kernels," *BMC Bioinformatics*, vol. 12, article 446, 2011.

[132] Y. Kim, A. Sette, and B. Peters, "Applications for T-cell epitope queries and tools in the Immune Epitope Database and Analysis Resource," *Journal of Immunological Methods*, vol. 374, no. 1-2, pp. 62–69, 2011.

[133] J. Ponomarenko, N. Papangelopoulos, D. M. Zajonc, B. Peters, A. Sette, and P. E. Bourne, "IEDB-3D: structural data within the immune epitope database," *Nucleic Acids Research*, vol. 39, no. 1, pp. D1164–D1170, 2011.

[134] S. Modrow, B. H. Hahn, and G. M. Shaw, "Computer-assisted analysis of envelope protein sequences of seven human immunodeficiency virus isolates: prediction of antigenic epitopes in conserved and variable regions," *Journal of Virology*, vol. 61, no. 2, pp. 570–578, 1987.

[135] J. R. A. Schafer, B. M. Jesdale, J. A. George, N. M. Kouttab, and A. S. De Groot, "Prediction of well-conserved HIV-1 ligands using a matrix-based algorithm, EpiMatrix," *Vaccine*, vol. 16, no. 19, pp. 1880–1884, 1998.

[136] M. H. Sung and R. Simon, "Genomewide conserved epitope profiles of HIV-1 predicted by biophysical properties of MHC binding peptides," *Journal of Computational Biology*, vol. 11, no. 1, pp. 125–145, 2004.

[137] Y. Xiao and M. R. Segal, "Prediction of genomewide conserved epitope profiles of HIV-1: classifier choice and peptide representation," *Statistical Applications in Genetics and Molecular Biology*, vol. 4, no. 1, article 25, 2005.

[138] M. Lacerda, K. Scheffler, and C. Seoighe, "Epitope discovery with phylogenetic hidden markov models," *Molecular Biology and Evolution*, vol. 27, no. 5, pp. 1212–1220, 2010.

[139] P. Rusert, A. Krarup, C. Magnus et al., "Interaction of the gp120 V1V2 loop with a neighboring gp120 unit shields the HIV envelope trimer against cross-neutralizing antibodies," *Journal of Experimental Medicine*, vol. 208, no. 7, pp. 1419–1433, 2011.

[140] B. E. Pickett, E. L. Sadat, Y. Zhang et al., "ViPR: an open bioinformatics database and analysis resource for virology research," *Nucleic Acids Research*, vol. 40, pp. D593–D598, 2012.

[141] X. Hu, W. Zhou, K. Udaka, H. Mamitsuka, and S. Zhu, "MetaMHC: a meta approach to predict peptides binding to MHC molecules," *Nucleic Acids Research*, vol. 38, no. 2, pp. W474–W479, 2010.

[142] N. C. Toussaint and O. Kohlbacher, "OptiTope—a web server for the selection of an optimal set of peptides for epitope-based vaccines," *Nucleic Acids Research*, vol. 37, no. 2, pp. W617–W622, 2009.

[143] M. Garcia-Boronat, C. M. Diez-Rivero, E. L. Reinherz, and P. A. Reche, "PVS: a web server for protein sequence variability analysis tuned to facilitate conserved epitope discovery," *Nucleic Acids Research*, vol. 36, pp. W35–W41, 2008.

[144] P. A. Reche and E. L. Reinherz, "PEPVAC: a web server for multi-epitope vaccine development based on the prediction of supertypic MHC ligands," *Nucleic Acids Research*, vol. 33, no. 2, pp. W138–W142, 2005.

[145] P. A. Reche, J. P. Glutting, and E. L. Reinherz, "Prediction of MHC class I binding peptides using profile motifs," *Human Immunology*, vol. 63, no. 9, pp. 701–709, 2002.

[146] P. A. Reche, J. P. Glutting, H. Zhang, and E. L. Reinherz, "Enhancement to the RANKPEP resource for the prediction of peptide binding to MHC molecules using profiles," *Immunogenetics*, vol. 56, no. 6, pp. 405–419, 2004.

[147] T. Lengauer and M. Rarey, "Computational methods for biomolecular docking," *Current Opinion in Structural Biology*, vol. 6, no. 3, pp. 402–406, 1996.

[148] C. Hansch, "Dihydrofolate reductase inhibition. A study in the use of X-ray crystallography, molecular graphics, and quantitative structure-activity relations in drug design," *Drug*

Intelligence and Clinical Pharmacy, vol. 16, no. 5, pp. 391–395, 1982.

[149] I. D. Kuntz, J. M. Blaney, S. J. Oatley, R. Langridge, and T. E. Ferrin, "A geometric approach to macromolecule-ligand interactions," *Journal of Molecular Biology*, vol. 161, no. 2, pp. 269–288, 1982.

[150] E. Horjales and C. I. Branden, "Docking of cyclohexanol-derivatives into the active site of liver alcohol dehydrogenase. Using computer graphics and energy minimization," *Journal of Biological Chemistry*, vol. 260, no. 29, pp. 15445–15451, 1985.

[151] R. L. DesJarlais, R. P. Sheridan, J. S. Dixon, I. D. Kuntz, and R. Venkataraghavan, "Docking flexible ligands to macromolecular receptors by molecular shape," *Journal of Medicinal Chemistry*, vol. 29, no. 11, pp. 2149–2153, 1986.

[152] F. M. L. G. Stamato, E. Longo, R. Ferreira, and O. Tapia, "The catalytic mechanism of serine proteases. III. An Indo-ISCRF study of the methylacetate docking in α-chymotrypsin," *Journal of Theoretical Biology*, vol. 118, no. 1, pp. 45–59, 1986.

[153] W. L. Jorgensen, "Rusting of the lock and key model for protein-ligand binding," *Science*, vol. 254, no. 5034, pp. 954–955, 1991.

[154] S. Chaudhury and J. J. Gray, "Conformer selection and induced fit in flexible backbone protein-protein docking using computational and NMR ensembles," *Journal of Molecular Biology*, vol. 381, no. 4, pp. 1068–1087, 2008.

[155] A. J. Olson and D. S. Goodsell, "Automated docking and the search for HIV protease inhibitors," *SAR and QSAR in environmental research*, vol. 8, no. 3-4, pp. 273–285, 1998.

[156] I. Muegge, Y. C. Martin, P. J. Hajduk, and S. W. Fesik, "Evaluation of PMF scoring in docking weak ligands to the FK506 binding protein," *Journal of Medicinal Chemistry*, vol. 42, no. 14, pp. 2498–2503, 1999.

[157] S. Lise, A. Walker-Taylor, and D. T. Jones, "Docking protein domains in contact space," *BMC Bioinformatics*, vol. 7, article 310, 2006.

[158] F. N. Novikov, V. S. Stroylov, O. V. Stroganov, V. Kulkov, and G. G. Chilov, "Developing novel approaches to improve binding energy estimation and virtual screening: a PARP case study," *Journal of Molecular Modeling*, vol. 15, no. 11, pp. 1337–1347, 2009.

[159] B. G. Pierce, Y. Hourai, and Z. Weng, "Accelerating protein docking in ZDOCK using an advanced 3D convolution library," *PLoS ONE*, vol. 6, no. 9, Article ID e24657, 2011.

[160] H. Cho, C. W. Yun, W. K. Park et al., "Modulation of the activity of pro-inflammatory enzymes, COX-2 and iNOS, by chrysin derivatives," *Pharmacological Research*, vol. 49, no. 1, pp. 37–43, 2004.

[161] S. Bai, T. Du, and E. Khosravi, "Applying internal coordinate mechanics to model the interactions between 8R-lipoxygenase and its substrate," *BMC Bioinformatics*, vol. 11, supplement 6, article S2, 2010.

[162] A. Pedretti, L. Villa, and G. Vistoli, "Modeling of binding modes and inhibition mechanism of some natural ligands of farnesyl transferase using molecular docking," *Journal of Medicinal Chemistry*, vol. 45, no. 7, pp. 1460–1465, 2002.

[163] N. K. U. Koehler, C. Y. Yang, J. Varady et al., "Structure-based discovery of nonpeptidic small organic compounds to block the T cell response to myelin basic protein," *Journal of Medicinal Chemistry*, vol. 47, no. 21, pp. 4989–4997, 2004.

[164] A. W. Michels, D. A. Ostrov, L. Zhang et al., "Structure-based selection of small molecules to alter allele-specific MHC class II antigen presentation," *Journal of Immunology*, vol. 187, no. 11, pp. 5921–5930, 2011.

[165] Zaheer-Ul-Haq and W. Khan, "Molecular and structural determinants of adamantyl susceptibility to HLA-DRs allelic variants: an in silico approach to understand the mechanism of MLEs," *Journal of Computer-Aided Molecular Design*, vol. 25, no. 1, pp. 81–101, 2011.

[166] A. Bhinge, P. Chakrabarti, K. Uthanumallian, K. Bajaj, K. Chakraborty, and R. Varadarajan, "Accurate detection of protein:ligand binding sites using molecular dynamics simulations," *Structure*, vol. 12, no. 11, pp. 1989–1999, 2004.

[167] X. N. Tang, C. W. Lo, Y. C. Chuang et al., "Prediction of the binding mode between GSK3β and a peptide derived from GSKIP using molecular dynamics simulation," *Biopolymers*, vol. 95, no. 7, pp. 461–471, 2011.

[168] X. Li, O. Keskin, B. Ma, R. Nussinov, and J. Liang, "Protein-protein interactions: hot spots and structurally conserved residues often locate in complemented pockets that preorganized in the unbound states: implications for docking," *Journal of Molecular Biology*, vol. 344, no. 3, pp. 781–795, 2004.

[169] B. Huang and M. Schroeder, "Using protein binding site prediction to improve protein docking," *Gene*, vol. 422, no. 1-2, pp. 14–21, 2008.

[170] C. Hetényi and D. van der Spoel, "Toward prediction of functional protein pockets using blind docking and pocket search algorithms," *Protein Science*, vol. 20, no. 5, pp. 880–893, 2011.

[171] S. Eyrisch and V. Helms, "Transient pockets on protein surfaces involved in protein-protein interaction," *Journal of Medicinal Chemistry*, vol. 50, no. 15, pp. 3457–3464, 2007.

[172] Y. Duan, B. V. B. Reddy, and Y. N. Kaznessis, "Residue conservation information for generating near-native structures in protein-protein docking," *Journal of Bioinformatics and Computational Biology*, vol. 4, no. 4, pp. 793–806, 2006.

[173] S. J. Cockell, B. Oliva, and R. M. Jackson, "Structure-based evaluation of in silico predictions of protein-protein interactions using comparative docking," *Bioinformatics*, vol. 23, no. 5, pp. 573–581, 2007.

[174] C. L. Pierri, G. Parisi, and V. Porcelli, "Computational approaches for protein function prediction: a combined strategy from multiple sequence alignment to molecular docking-based virtual screening," *Biochimica et Biophysica Acta*, vol. 1804, no. 9, pp. 1695–1712, 2010.

[175] G. Heberlé and W. F. de Azevedo, "Bio-inspired algorithms applied to molecular docking simulations," *Current Medicinal Chemistry*, vol. 18, no. 9, pp. 1339–1352, 2011.

[176] P. N. Palma, L. Krippahl, J. E. Wampler, and J. J. G. Moura, "Bigger: a new (soft) docking algorithm for predicting protein interactions," *Proteins*, vol. 39, no. 4, pp. 372–384, 2000.

[177] F. Giordanetto, S. Cotesta, C. Catana et al., "Novel scoring functions comprising QXP, SASA, and protein side-chain entropy terms," *Journal of Chemical Information and Computer Sciences*, vol. 44, no. 3, pp. 882–893, 2004.

[178] A. D. J. van Dijk, S. J. De Vries, C. Dominguez, H. Chen, H.-X. Zhou, and A. M. J. J. Bonvin, "Data-driven docking: HADDOCK'S adventures in CAPRI," *Proteins*, vol. 60, no. 2, pp. 232–238, 2005.

[179] S. Betzi, K. Suhre, B. Chétrit, F. Guerlesquin, and X. Morelli, "GFscore: a general nonlinear consensus scoring function for high-throughput docking," *Journal of Chemical Information and Modeling*, vol. 46, no. 4, pp. 1704–1712, 2006.

[180] J. D. Durrant and J. A. McCammon, "NNScore: a neural-network-based scoring function for the characterization of

protein-ligand complexes," *Journal of Chemical Information and Modeling*, vol. 50, no. 10, pp. 1865–1871, 2010.

[181] J. D. Durrant and J. A. McCammon, "NNScore 2.0: a neural-network receptor-ligand scoring function," *Journal of Chemical Information and Modeling*, vol. 51, no. 11, pp. 2897–2903, 2011.

[182] S. Ahmad and K. Mizuguchi, "Partner-aware prediction of interacting residues in protein-protein complexes from sequence data," *PLoS ONE*, vol. 6, no. 12, Article ID e29104, 2011.

[183] A. Grosdidier, V. Zoete, and O. Michielin, "SwissDock, a protein-small molecule docking web service based on EADock DSS," *Nucleic Acids Research*, vol. 39, no. 2, pp. W270–W277, 2011.

[184] E. Dolghih, C. Bryant, A. R. Renslo, and M. P. Jacobson, "Predicting binding to P-glycoprotein by flexible receptor docking," *PLoS Computational Biology*, vol. 7, no. 6, Article ID e1002083, 2011.

[185] N. D. Prakhov, A. L. Chernorudskiy, and M. R. Gainullin, "VSDocker: a tool for parallel high-throughput virtual screening using AutoDock on Windows-based computer clusters," *Bioinformatics*, vol. 26, no. 10, pp. 1374–1375, 2010.

[186] J. Ren, N. Williams, L. Clementi, S. Krishnan, and W. W. Li, "Opal web services for biomedical applications," *Nucleic Acids Research*, vol. 38, no. 2, pp. W724–W731, 2010.

[187] Q. Zhang, J. Yang, K. Liang et al., "Binding interaction analysis of the active site and its inhibitors for neuraminidase (N1 Subtype) of human influenza virus by the integration of molecular ducking, FMO calculation and 3D-QSAR CoMFA modeling," *Journal of Chemical Information and Modeling*, vol. 48, no. 9, pp. 1802–1812, 2008.

[188] H. Li, Z. Gao, L. Kang et al., "TarFisDock: a web server for identifying drug targets with docking approach," *Nucleic Acids Research*, vol. 34, pp. W219–W224, 2006.

[189] L. Martin, V. Catherinot, and G. Labesse, "kinDOCK: a tool for comparative docking of protein kinase ligands," *Nucleic Acids Research*, vol. 34, pp. W325–W329, 2006.

[190] A. Lavecchia, S. Cosconati, V. Limongelli, and E. Novellino, "Modeling of Cdc25B dual specifity protein phosphatase inhibitors: docking of ligands and enzymatic inhibition mechanism," *ChemMedChem*, vol. 1, no. 5, pp. 540–550, 2006.

[191] M. Schapira, B. M. Raaka, S. Das et al., "Discovery of diverse thyroid hormone receptor antagonists by high-throughput docking," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 12, pp. 7354–7359, 2003.

[192] R. Chen, L. Li, and Z. Weng, "ZDOCK: an initial-stage protein-docking algorithm," *Proteins*, vol. 52, no. 1, pp. 80–87, 2003.

[193] B. Pierce, W. Tong, and Z. Weng, "M-ZDOCK: a grid-based approach for Cn symmetric multimer docking," *Bioinformatics*, vol. 21, no. 8, pp. 1472–1478, 2005.

[194] A. W. Ghoorah, M. D. Devignes, M. Smaïl-Tabbone, and D. W. Ritchie, "Spatial clustering of protein binding sites for template based protein docking," *Bioinformatics*, vol. 27, no. 20, pp. 2820–2827, 2011.

[195] S. Lyskov and J. J. Gray, "The RosettaDock server for local protein-protein docking," *Nucleic acids research*, vol. 36, pp. W233–238, 2008.

[196] S. R. Comeau, D. W. Gatchell, S. Vajda, and C. J. Camacho, "ClusPro: an automated docking and discrimination method

for the prediction of protein complexes," *Bioinformatics*, vol. 20, no. 1, pp. 45–50, 2004.

[197] S. R. Comeau, D. W. Gatchell, S. Vajda, and C. J. Camacho, "ClusPro: a fully automated algorithm for protein-protein docking," *Nucleic Acids Research*, vol. 32, pp. W96–W99, 2004.

[198] I. Tuszynska and J. M. Bujnicki, "DARS-RNP and QUASI-RNP: new statistical potentials for protein-RNA docking," *BMC Bioinformatics*, vol. 12, article 348, 2011.

[199] Z. Liu, J. T. Guo, T. Li, and Y. Xu, "Structure-based prediction of transcription factor binding sites using a protein-DNA docking approach," *Proteins*, vol. 72, no. 4, pp. 1114–1124, 2008.

[200] B. Xu, Y. Yang, H. Liang, and Y. Zhou, "An all-atom knowledge-based energy function for protein-DNA threading, docking decoy discrimination, and prediction of transcription-factor binding profiles," *Proteins*, vol. 76, no. 3, pp. 718–730, 2009.

[201] A. Patronov, I. Dimitrov, D. R. Flower, and I. Doytchinova, "Peptide binding prediction for the human class II MHC allele HLA-DP2: a molecular docking approach," *BMC Structural Biology*, vol. 11, article 32, 2011.

[202] A. Sircar and J. J. Gray, "SnugDock: paratope structural optimization during antibody-antigen docking compensates for errors in antibody homology models," *PLoS Computational Biology*, vol. 6, no. 1, Article ID e1000644, 2010.

[203] J. M. Khan and S. Ranganathan, "PDOCK: a new technique for rapid and accurate docking of peptide ligands to Major Histocompatibility Complexes," *Immunome Research*, vol. 6, supplement 1, article S2, 2010.

[204] S. J. Todman, M. D. Halling-Brown, M. N. Davies, D. R. Flower, M. Kayikci, and D. S. Moss, "Toward the atomistic simulation of T cell epitopes. Automated construction of MHC: peptide structures for free energy calculations," *Journal of Molecular Graphics and Modelling*, vol. 26, no. 6, pp. 957–961, 2008.

[205] R. Sathyapriya, M. S. Vijayabaskar, and S. Vishveshwara, "Insights into protein-DNA interactions through structure network analysis," *PLoS Computational Biology*, vol. 4, no. 9, Article ID e1000170, 2008.

[206] L. Saiz and J. M. G. Vilar, "Protein-protein/DNA interaction networks: versatile macromolecular structures for the control of gene expression," *IET Systems Biology*, vol. 2, no. 5, pp. 247–255, 2008.

[207] J. Mata, "Genome-wide mapping of myosin protein-RNA networks suggests the existence of specialized protein production sites," *The FASEB Journal*, vol. 24, no. 2, pp. 479–484, 2010.

[208] M. Ascano, M. Hafner, P. Cekan, S. Gerstberger, and T. Tuschl, "Identification of RNA-protein interaction networks using PAR-CLIP," *Wiley Interdisciplinary Reviews: RNA*, vol. 3, no. 2, pp. 159–177, 2011.

[209] V. Spirin and L. A. Mirny, "Protein complexes and functional modules in molecular networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 21, pp. 12123–12128, 2003.

[210] J. Wang, M. Li, Y. Deng, and Y. Pan, "Recent advances in clustering methods for protein interaction networks," *BMC Genomics*, vol. 11, no. 3, article S10, 2010.

[211] Z. Wu, X. Zhao, and L. Chen, "Identifying responsive functional modules from protein-protein interaction network," *Molecules and Cells*, vol. 27, no. 3, pp. 271–277, 2009.

[212] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, "Network motifs: simple building blocks of

complex networks," *Science*, vol. 298, no. 5594, pp. 824–827, 2002.

[213] G. Ciriello and C. Guerra, "A review on models and algorithms for motif discovery in protein-protein interaction networks," *Briefings in Functional Genomics and Proteomics*, vol. 7, no. 2, pp. 147–156, 2008.

[214] J. Dutkowski and J. Tiuryn, "Identification of functional modules from conserved ancestral protein-protein interactions," *Bioinformatics*, vol. 23, no. 13, pp. i149–i158, 2007.

[215] X. Qian and B. J. Yoon, "Effective identification of conserved pathways in biological networks using hidden Markov models," *PLoS ONE*, vol. 4, no. 12, Article ID e8070, 2009.

[216] J. Lim, T. Hao, C. Shaw et al., "A protein-protein interaction network for human inherited ataxias and disorders of purkinje cell degeneration," *Cell*, vol. 125, no. 4, pp. 801–814, 2006.

[217] L. Wang, P. Khankhanian, S. E. Baranzini, and P. Mousavi, "iCTNet: a Cytoscape plugin to produce and analyze integrative complex traits networks," *BMC Bioinformatics*, vol. 12, article 380, 2011.

[218] S. Suthram, J. T. Dudley, A. P. Chiang, R. Chen, T. J. Hastie, and A. J. Butte, "Network-based elucidation of human disease similarities reveals common functional modules enriched for pluripotent drug targets," *PLoS Computational Biology*, vol. 6, no. 2, Article ID e1000662, 2010.

[219] B. Dutta, L. Pusztai, Y. Qi et al., "A network-based, integrative study to identify core biological pathways that drive breast cancer clinical subtypes," *British Journal of Cancer*, vol. 106, no. 6, pp. 1107–1116, 2012.

[220] S. Erten, G. Bebek, and M. Koyutürk, "Vavien: an algorithm for prioritizing candidate disease genes based on topological similarity of proteins in interaction networks," *Journal of Computational Biology*, vol. 18, no. 11, pp. 1561–1574, 2011.

[221] D. Nitsch, L. C. Tranchevent, J. P. Gonalves, J. K. Vogt, S. C. Madeira, and Y. Moreau, "PINTA: a web server for network-based gene prioritization from expression data," *Nucleic Acids Research*, vol. 39, no. 2, pp. W334–W338, 2011.

[222] A. K. Dunker, M. S. Cortese, P. Romero, L. M. Iakoucheva, and V. N. Uversky, "Flexible nets: the roles of intrinsic disorder in protein interaction networks," *The FEBS Journal*, vol. 272, no. 20, pp. 5129–5148, 2005.

[223] A. Patil, K. Kinoshita, and H. Nakamura, "Hub promiscuity in protein-protein interaction networks," *International Journal of Molecular Sciences*, vol. 11, no. 4, pp. 1930–1943, 2010.

[224] A. Patil, K. Kinoshita, and H. Nakamura, "Domain distribution and intrinsic disorder in hubs in the human protein-protein interaction network," *Protein Science*, vol. 19, no. 8, pp. 1461–1468, 2010.

[225] D. M. Bustos, "The role of protein disorder in the 14-3-3 interaction network," *Molecular Biosystems*, vol. 8, no. 1, pp. 178–184, 2012.

[226] R. Ivanyi-Nagy and J. L. Darlix, "Intrinsic disorder in the core of flaviviruses," *Protein and Peptide Letters*, vol. 17, no. 8, pp. 1019–1025, 2010.

[227] J. Wang, Z. Cao, L. Zhao, and S. Li, "Novel strategies for drug discovery based on intrinsically disordered proteins (IDPs)," *International Journal of Molecular Sciences*, vol. 12, no. 5, pp. 3205–3219, 2011.

[228] D. Ekman, S. Light, A. K. Björklund, and A. Elofsson, "What properties characterize the hub proteins of the protein-protein interaction network of Saccharomyces cerevisiae?" *Genome Biology*, vol. 7, no. 6, article R45, 2006.

[229] G. P. Singh, M. Ganapathi, and D. Dash, "Role of intrinsic disorder in transient interactions of hub proteins," *Proteins*, vol. 66, no. 4, pp. 761–765, 2007.

[230] B. Kahali, S. Ahmad, and T. C. Ghosh, "Exploring the evolutionary rate differences of party hub and date hub proteins in Saccharomyces cerevisiae protein-protein interaction network," *Gene*, vol. 429, no. 1-2, pp. 18–22, 2009.

[231] A. R. Katebi, A. Kloczkowski, and R. L. Jernigan, "Structural interpretation of protein-protein interaction network," *BMC Structural Biology*, vol. 10, supplement 1, article S4, 2010.

[232] F. Wang, M. Liu, B. Song et al., "Prediction and characterization of protein-protein interaction networks in swine," *Proteome Science*, vol. 10, article 2, 2012.

[233] E. Pang and K. Lin, "Yeast protein-protein interaction binding sites: prediction from the motif-motif, motif-domain and domain-domain levels," *Molecular BioSystems*, vol. 6, no. 11, pp. 2164–2173, 2010.

[234] M. Kalaev, M. Smoot, T. Ideker, and R. Sharan, "NetworkBLAST: comparative analysis of protein networks," *Bioinformatics*, vol. 24, no. 4, pp. 594–596, 2008.

[235] S. H. Jung, B. Hyun, W. H. Jang, H. Y. Hur, and D. S. Han, "Protein complex prediction based on simultaneous protein interaction network," *Bioinformatics*, vol. 26, no. 3, pp. 385–391, 2009.

[236] Y.-S. Lo, C. Y. Lin, and J. M. Yang, "PCFamily: a web server for searching homologous protein complexes," *Nucleic Acids Research*, vol. 38, no. 2, pp. W516–W522, 2010.

[237] H. T. Phan and M. J. Sternberg, "PINALOG: a novel approach to align protein interaction networks—implications for complex detection and function prediction," *Bioinformatics*, vol. 28, pp. 1239–1245, 2012.

[238] A. D. Fox, B. J. Hescott, A. C. Blumer, and D. K. Slonim, "Connectedness of PPI network neighborhoods identifies regulatory hub proteins," *Bioinformatics*, vol. 27, no. 8, pp. 1135–1142, 2011.

[239] E. Becker, B. Robisson, C. E. Chapple, A. Guénoche, and C. Brun, "Multifunctional proteins revealed by overlapping clustering in protein interaction network," *Bioinformatics*, vol. 28, no. 1, pp. 84–90, 2012.

[240] M. T. Dittrich, G. W. Klau, A. Rosenwald, T. Dandekar, and T. Müller, "Identifying functional modules in protein-protein interaction networks: an integrated exact approach," *Bioinformatics*, vol. 24, no. 13, pp. i223–i231, 2008.

[241] C. H. Sun, T. Hwang, K. Oh, and G. S. Yi, "DynaMod: dynamic functional modularity analysis," *Nucleic Acids Research*, vol. 38, no. 2, pp. W103–W108, 2010.

[242] G. Cui, R. Shrestha, and K. Han, "ModuleSearch: finding functionalmodules in a protein-protein interaction network," *Computer Methods in Biomechanics and Biomedical Engineering*, vol. 15, no. 7, pp. 691–699, 2012.

[243] S. Erten, G. Bebek, R. M. Ewing, and M. Koyutürk, "DADA: degree-aware algorithms for network-based disease gene prioritization," *BioData Mining*, vol. 4, no. 1, article 19, 2011.

[244] P. Shannon, A. Markiel, O. Ozier et al., "Cytoscape: a software environment for integrated models of biomolecular interaction networks," *Genome Research*, vol. 13, no. 11, pp. 2498–2504, 2003.

[245] A. Martin, M. E. Ochagavia, L. C. Rabasa, J. Miranda, J. Fernandez-de-Cossio, and R. Bringas, "BisoGenet: a new tool for gene network building, visualization and analysis," *BMC Bioinformatics*, vol. 11, article 91, 2010.

[246] S. Bruckner, F. Hüffner, R. M. Karp, R. Shamir, and R. Sharan, "Torque: topology-free querying of protein interaction

networks," *Nucleic Acids Research*, vol. 37, no. 2, pp. W106–W108, 2009.

[247] E. Franzosa, B. Linghu, and Y. Xia, "Computational reconstruction of protein-protein interaction networks: algorithms and issues," *Methods in Molecular Biology*, vol. 541, pp. 89–100, 2009.

[248] S. S. Li and C. Wu, "Using peptide array to identify binding motifs and interaction networks for modular domains," *Methods in Molecular Biology*, vol. 570, pp. 67–76, 2009.

[249] M. E. Sardiu and M. P. Washburn, "Building protein-protein interaction networks with proteomics and informatics tools," *Journal of Biological Chemistry*, vol. 286, no. 27, pp. 23645–23651, 2011.

[250] R. Sanz-Pamplona, A. Berenguer, X. Sole et al., "Tools for protein-protein interaction network analysis in cancer research," *Clinical and Translational Oncology*, vol. 14, no. 1, pp. 3–14, 2012.

[251] D. Szklarczyk, A. Franceschini, M. Kuhn et al., "The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored," *Nucleic Acids Research*, vol. 39, no. 1, pp. D561–D568, 2011.

[252] O. Taboureau, S. K. Nielsen, K. Audouze et al., "ChemProt: a disease chemical biology database," *Nucleic Acids Research*, vol. 39, no. 1, pp. D367–D372, 2011.

[253] R. Tacutu, A. Budovsky, and V. E. Fraifeld, "The NetAge database: a compendium of networks for longevity, age-related diseases and associated processes," *Biogerontology*, vol. 11, no. 4, pp. 513–522, 2010.

[254] B.-J. Breitkreutz, C. Stark, T. Reguly et al., "The BioGRID interaction Database: 2008 update," *Nucleic Acids Research*, vol. 36, no. 1, pp. D637–D640, 2008.

[255] C. N. Hsu, J. M. Lai, C. H. Liu et al., "Detection of the inferred interaction network in hepatocellular carcinoma from EHCO (Encyclopedia of Hepatocellular Carcinoma genes Online)," *BMC Bioinformatics*, vol. 8, article 66, 2007.

[256] K. Xia, D. Dong, and J. D. J. Han, "IntNetDB v1.0: an integrated protein-protein interaction network database generated by a probabilistic model," *BMC Bioinformatics*, vol. 7, article 508, 2006.

[257] J. H. Fong and A. Marchler-Bauer, "Protein subfamily assignment using the Conserved Domain Database," *BMC Research Notes*, vol. 1, article 114, 2008.

[258] O. V. Stroganov, F. N. Novikov, V. S. Stroylov, V. Kulkov, and G. G. Chilov, "Lead finder: an approach to improve accuracy of protein-ligand docking, binding energy estimation, and virtual screening," *Journal of Chemical Information and Modeling*, vol. 48, no. 12, pp. 2371–2385, 2008.