New Developments in Protein Structure Modelling for Biological and Clinical Research

InteractoMIX: a suite of computational tools to exploit interactomes in biological and clinical research

Daniel Poglayen*[1], Manuel Alejandro Marín-López*[1], Jaume Bonet*†[1], Oriol Fornes*‡, Javier Garcia-Garcia*, Joan Planas-Iglesias§, Joan Segura‖, Baldo Oliva*[2] and Narcis Fernandez-Fuentes¶[2]

*Structural Bioinformatics Lab (GRIB-IMIM), Department of Experimental and Health Science, University Pompeu Fabra, Barcelona 08005, Catalonia, Spain

†Laboratory of Protein Design & Immunoengineering, School of Engineering, Ecole Polytechnique Federale de Lausanne, Lausanne 1015, Vaud, Switzerland

‡Centre for Molecular Medicine and Therapeutics at the Child and Family Research Institute, Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada V5Z 4H4

§Division of Metabolic and Vascular Health, University of Warwick, Coventry CV4 7AL, U.K.

‖GN7 of the National Institute for Bioinformatics (INB) and Biocomputing Unit, National Center of Biotechnology (CSIC), Madrid 28049, Spain

¶Institute of Biological, Environmental and Rural Sciences, Aberystwyth University, Aberystwyth SY23 3DA, U.K.

Abstract

Virtually all the biological processes that occur inside or outside cells are mediated by protein–protein interactions (PPIs). Hence, the charting and description of the PPI network, initially in organisms, the interactome, but more recently in specific tissues, is essential to fully understand cellular processes both in health and disease. The study of PPIs is also at the heart of renewed efforts in the medical and biotechnological arena in the quest of new therapeutic targets and drugs. Here, we present a mini review of 11 computational tools and resources tools developed by us to address different aspects of PPIs: from interactome level to their atomic 3D structural details. We provided details on each specific resource, aims and purpose and compare with equivalent tools in the literature. All the tools are presented in a centralized, one-stop, web site: InteractoMIX (http://interactomix.com).

Introduction

The technological advances in high-throughput technologies, heralded by the human genome sequencing project [1], represented a radical change resulting in an exponential increase in biological data and the birth of the different omic technologies that ushered life sciences in the post-genomic era. The formidable wealth of information on gene and protein sequences, protein structures and functions (both individually and collectively), have led to new challenges related to handling and, more importantly, deciphering and understanding this information. Protein–protein interactions (PPIs) are crucial for understanding how cells work, and yet, even for well-studied organisms only approximately 42% of their interactomes have been described [2]. Thus computational tools, in combination with experimental data, have an important role to play towards a better comprehension of function and behaviour of proteins in biological systems.

The function of a protein can be inferred from its interactions with other proteins by means of the '*guilt by association principle*': proteins involved in the same biological process tend to cluster together in PPI networks and, in turn, the paths derived from these networks increase the knowledge on the process itself [3]. PPI sub-networks involved in fundamental processes are extremely coordinated

Key words: computational predictions, databases, protein–protein interactions, structural bioinformatics, web-servers.

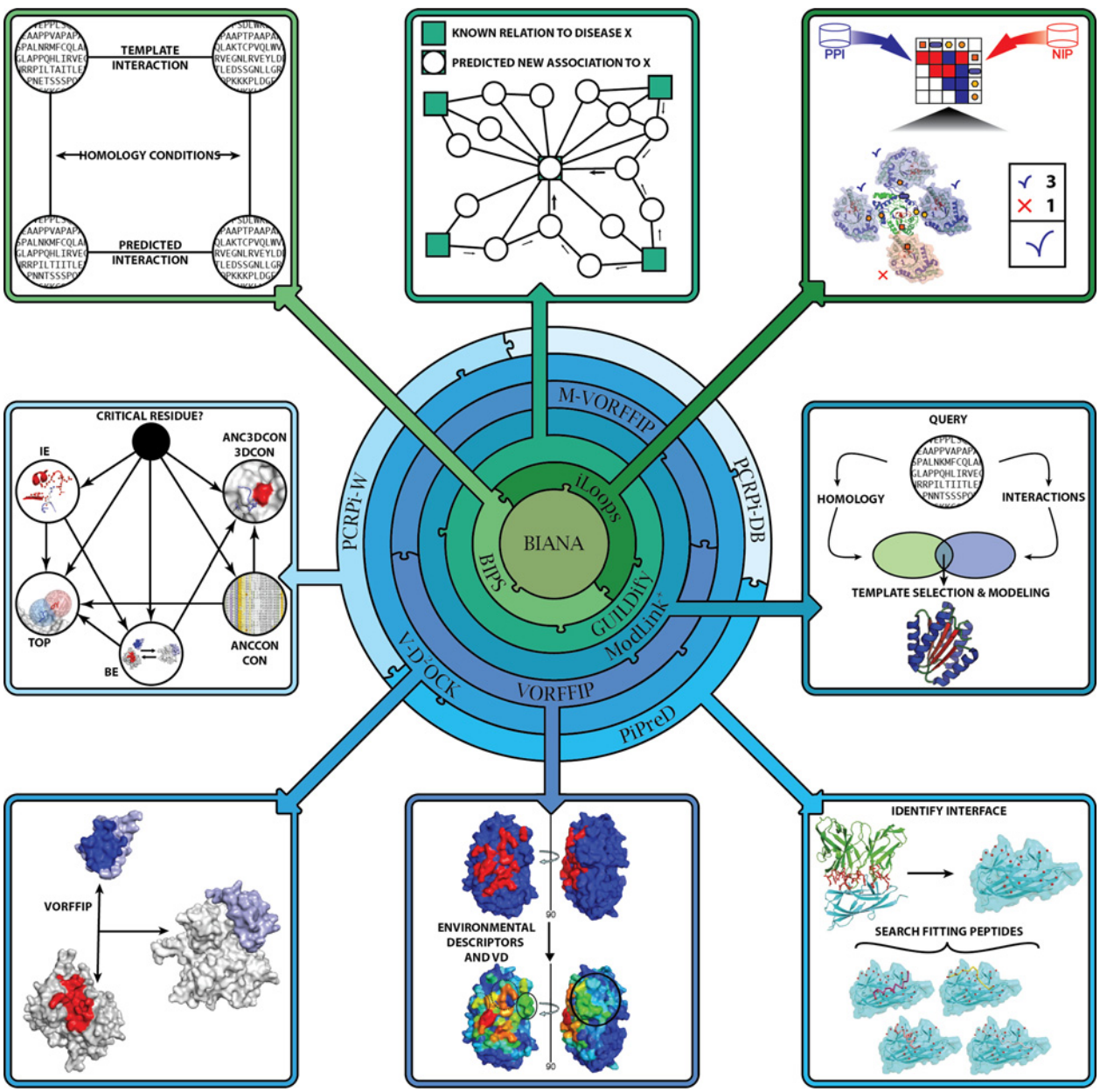
Abbreviations: BIANA, biologic interactions and network analysis; NIP, non-interacting proteins; PPI, protein–protein interaction.

[1]**These authors contributed equally to this work.**

[2]**Correspondence may be addressed to either of these authors (email baldo.oliva@upf.edu or narcis.fernandez@gmail.com).**

Figure 1 | Diagrammatic representation of the set of integrated tools in InteractoMIX

A bull's eye representation in the centre of the image represents the integration of the tools in InteractoMIX: tools for interactomic-level analyses are represented in shades of green; tools for atomic-level analyses are represented in shades of blue. The puzzle-like connections represent the putative dependency between the different tools, thus indicating a workflow that can be followed along a particular study involving PPIs. In the different insets from the top left side and clock-wise schematic representation of different methods: BIPS: prediction of protein interactions based in interologs; GUILDify: network-based methods to uncover novel disease-related genes; iLoops: prediction of protein structures based on local motifs; ModLink+: network-based selection of protein templates for structure prediction; PiPreD: modelling of orthosteric peptides; VORFFIP and M-VORFFIP: prediction of functional sites in protein structures; VD^{2}OCK: data-driven protein docking and PCRPi: prediction of critical or hot spot residues in protein interfaces.



and highly complex, and they are the basis of the even more complicated cell machinery. Thus, describing all existing interactions in an organism, i.e. its interactome, is central in different aspects of modern research in biomedicine and biotechnology. Furthermore, PPIs are key in the new efforts directed towards the discovery of new therapeutic targets and drugs and for the development of new promising fields such as personalized and preventive medicine [4,5].

Here we present a mini-review of 11 published tools and databases for the analysis of PPIs, ranging from the

interactome level to their structural details. These tools are been centralized is a web site: InteractoMIX, designed to facilitate the access to these resources in an easy and intuitive manner. As depicted in Figure 1, InteractoMIX addresses two major aspects of PPIs: (1) what proteins interact and (2) how do they interact (i.e. the structural details of the PPI). The aims and specific use(s) of the tools are described below and in more detail in each of sections devoted to each of them.

At a high-level, InteractoMIX includes tools to integrate interactomic data, predicting new interactions and genotype-phenotype associations. The first step is to use a specific network of PPIs. Several databases contain sources of PPIs (i.e. IntAct [6], DIP [7], BioGRID [8], HPRD [9], MINT [10], MPact [11] among others). Nevertheless, one of the main problems is the integration of these sources. Similarly to approaches such as iRefIndex [12] or APID [13] we suggest the use of BIANA (biologic interactions and network analysis) [14], in which the user controls the integration. Using BIANA has the advantage to integrate functional information from other sources, such as Uniprot [15], PFAM [16] or GO [17] which can further be used in the server GUIDify [18]. GUILDify uses the principle of 'guilt-by-association' to infer and prioritize a list of genes and proteins associated with a function or phenotype. The approach uses several algorithms described in the package GUILD [19] and a list of starting seeds (genes associated with a function or phenotype) extracted with the post-integration of BIANA. This approach has been used to reduce the human interactome to a selected set of relevant interactions in the study of cancer metastasis to lung and brain [20] or extend the subnetwork associated with the pathways of apoptosis [21]. Between both levels of prediction the user can increase the network with predicted interactions. Two different approaches are accessible in InteractoMIX: the first one uses the integration of PPIs and protein sequences to predict interologs (i.e. the interaction between two proteins which sequence is sufficiently similar to those of a known interacting pair); the second analyses domains and super-secondary structural motifs involved in interactions and assigns them to two query proteins to predict their potential interaction. The first approach uses the server BIPS [22] and has been very useful to predict networks of interactions between proteins of different species such as human and other pathogens (i.e. *Salmonella* [23] or *Mycobacterium* [24]), whereas the second uses iLoops [25,26] and it helps on the structural detection of interactions, such as the modelling of GR interactions during chromatin remodelling [27] or the complexes of proteins that interact with cardiolipin [28].

From a structural perspective, InteractoMIX includes tools for the modelling of the 3D structures of proteins and protein complexes and to characterize protein interfaces including the effect of mutations or post-transcriptional modifications including the modelling and design of peptides to target PPIs. Modlink + [29] is the first of the tools included in the structural analyses featured in InteractoMIX. It combines remote homology and network analysis to improve the selection of suitable templates for protein structure prediction particularly at low levels of sequence identity. VORFFIP [30] and M-VORFFIP [31] represents two tools designed to predict functional sites in protein structures, i.e. protein-, peptide-, DNA- and RNA-binding sites. Whether the structure of cognate partners is known but not as complex, VD^{2}OCK [32] is the tool of choice to derive structural models of the protein complex. VD^{2}OCK [32] makes use of VORFFIP [31] to delineate the protein interfaces and direct the docking sampling. The energy of associations in PPIs is dominated by few interactions, i.e. hot spots of interaction [33], and in this aspect PCRPi [34] predicts such critical interactions. The final tool included in structure-based set of InteractoMIX is PiPred [35] developed to predict orthosteric peptides with the view of modulating PPIs.

In summary, InteractoMIX provides a unique resource with set of functionalities devoted to the computational study of PPIs in a single, one stop, on line resource.

Exploiting the interactome: integration and unification of interactomic data

The dramatic increase in biological information witnessed over the post-genomic era has inevitably led to a number of challenges, one of these being the integration of biological databases handling interactomic data. Individual repositories use their unique nomenclature, storage systems and formats and given the large amount information and incredibly fast rate of generation, manual curation is an unpractical and often inefficient task. In addition, the increase in throughput in data generation has also resulted in the growth and propagation of unavoidable human and experimental errors. All these aspects are major obstacles in data integration and interpretation. Furthermore, unification is a non-trivial problem, and thus an effective integration strategy is key for the successful merge of multiple databases while increasing data completeness.

At the core of InteractoMIX is BIANA, last updated on February 2015, a platform designed to compile and integrate interactomic data from multiple sources in a comprehensive and traceable manner [14]. BIANA uses a high-level abstraction schema to integrate and define external repositories or databases compiling interactomics and protein–ligand (small chemical) information. The two unique features of BIANA, missing in other resources devoted to the same end (e.g. APID [13] or iRefIndex [12]) are: (i) its unification protocol (i.e. a set of rules defined by the users that determine how data from various sources is combined) offering also the possibility of cross-checking data across different databases and (ii) its traceability: merged entities can always be traced back to its original source.

BIANA can be used as a standalone application, through a dedicated online server or as a Cytoscape [36] plugin. Furthermore, although most integration approaches generate new datasets and their update depends on the authors, the update of the integration resulting from BIANA standalone software is in the hands of the user who can decided which subsets of the interactome and which databases are used upon releases of new information. BIANA conforms the foundations of the following tools included in InteractoMIX:

BIPS [22], an interolog-based prediction server; the network relationships used to classify structural features used in iLoops [26]; and network-based gene-disease associations exploited by GUILDify [18].

Exploiting the interactome: network-based tools

InteractoMIX includes two PPIs prediction methods that expand current interactomic knowledge from experimental databases (depicted as a range of green shades in Figure 1). On the one hand, BIPS infers PPIs based on the interologs approach [22]. On the other hand, iLoops predicts pairs of PPIs evaluating different structural features [25]. Finally, InteractoMIX incorporates GUILDify, a guild-by-association algorithm to predict gene–phenotype associations [18]. Potential applications of these tools include the increase in the coverage of PPIs networks, and the search for new therapeutic targets and their interactions.

BIPS: interolog-based PPIs predictions

BIPS [22] is an interolog-based PPIs prediction method that is based on the interolog hypothesis, which implies that two proteins (A and B) are predicted to interact if a known interaction between two homologue proteins (A′ and B′) exists [37]. The completeness of the database of PPIs and the option to run large queries are two central aspects of web-servers delivering such predictions. Accordingly, BIPS both benefits from the comprehensive interactomic data compiled in BIANA [14] and computes proteome-wide interactomes in a reasonable time, thanks to the usage of a local database of pre-computed similarity measures.

The input to BIPS web-server consists of a list of query proteins in either FASTA format or protein identifiers (e.g. Uniprot [38] codes) and outputs a list of predicted protein pairs. Users can inspect and download details relevant to the predictions (e.g. interolog mapping). Several parameters can be tuned to improve the reliability of the predictions, such as sequence similarity thresholds, number of experiments or species confirming the mapping or even the type of experiment describing such interactions. Furthermore, filters including known domain–domain interactions, functional similarity, taxonomy, associated pathologies or others can also be applied. BIPS has been widely used, to extend the signalling network in apoptosis [21] or in the comparison of *Salmonella*–hosts interactomes [23]. Similarly to BIPS, other approaches have further developed more sophisticated approaches to improve the accuracy of the prediction [39, 40], however one of the main advantages of BIPS which is the capacity to predict full interactomes or cross-species interactomes has not been overcome.

iLoops: local-based PPIs predictions

iLoops [25] is a computational tool designed to infer new and describe the molecular mechanisms of PPIs. iLoops is based on the evaluation of local structural features defining characteristic patterns of interaction (interaction signatures) learned from known PPIs and non-interacting proteins (NIP) pairs. A unique feature to iLoops is the classification of the interaction signatures as favouring or disfavouring depending on their role in facilitating or preventing the interaction between given pairs of proteins. By avoiding global sequence similarity, the method is able to evaluate the interaction of designed proteins, going far beyond the organisms' interactomes.

The input to iLoops is a set of FASTA sequences and a list of protein pairs to be evaluated and outputs a list of PPIs and NIP pairs. Details of each prediction can be explored in a visual interface or locally by downloading an XML file. iLoops web server is frequently visited worldwide averaging over 15 unique visits per month and has been already used in several projects such as the work by Na et al. [41] to elucidate whether Sup35pC can physically interact with actin/Ssa1/PIK3R1, and thus yielding insights into the role of Sup35pC as a cytoskeleton modulator. iLoops has been recently updated to include the novel structural motifs classified in ArchDB 2014 [42] resulting in an increase in the coverage of local motifs and thus increasing its applicability. iLoops has many advantages over other methods of prediction: first, it gives the user the probability of the prediction depending on the level of expectation suggested by the user, ranging between the simple random criterion with a 1/50 ratio (one positive interaction over 50 possible pairs is the expected ratio for co-localized proteins of the human proteome); and second, it only requires local assignation of structure instead of full domains (such as PRISM [43] or Interactome3D [44]), despite reducing its accuracy.

GUILDify: network-based disease gene-prioritization predictions

GUILDify [18] is a network-based tool to predict and rank genes linked to biological processes and disease phenotypes by combining experimental data and graph-based *guilt-by-association* algorithms. For a given phenotype and a set of core genes (e.g. known disease-genes or user-defined genes), GUILDify predicts and ranks potential novel genes associated with the given phenotype based on the connectedness to a core set of genes through the underlying PPIs network. Thus, it uses BIANA [14] to create an integrated knowledge base of protein-coding genes, their functional and disease annotations and the interactions between them.

In contrast with several existing network-based prioritization tools relying on user-defined gene sets (e.g. [45]) or annotations based on OMIM database (e.g. [46]), GUILDify automatically retrieves phenotype–gene associations and does not restrict the search to disease phenotypes. The genes matching to the user-provided query (e.g. 'type 2 diabetes') are then fed to GUILD [19], a network-based prioritization algorithm that assigns scores to the genes based on their distance to the genes matching the query. GUILDify has been designed for both experimental and computational biomedical researchers seeking to find an initial set of genes and a ranked list of genes potentially linked to the phenotype of interest. GUILDify has been used to identify genes

involved in bone metastatic breast cancer [47] or to predict novel uses of drugs (see Supplementary data on Guney et al. [18] for further details). Other similar approaches have been developed to prioritize genes associated with functions or patho-phenotypes (i.e. GeneMania [48] or DIAMOnD [49] are well known). Recently the method was used to test the resilience of certain phenotypes to be affected by the deletion of nodes on the basis of the robustness of the prediction [50,51].

Exploiting the interactome: structural-based tools

Included in this review are several resources compiled in InteractoMIX designed to depict the structural details of PPIs at a molecular level (represented in a range of blues shades in Figure 1). ModLink + is a tool to model the 3D-structure of individual proteins [29]. VORFFIP and M-VORFFIP [30,31] tools predict functional sites in proteins including protein interfaces, which can be used to guide V-D^{2}OCK, a fast and accurate docking programme [32]. From the selected interfaces, PCRPi-W elucidates the critical residues to the given interactions (i.e. hot-spot) [52], which are compiled in a weekly-updated database: PCRPi-DB [53]. Finally, PiPreD allows the modelling of orthosteric peptides to target and disrupt PPIs [12,35]. Potential applications of these tools are: identification of key residues in PPIs, designing protein mutants able to modulate PPIs and computational designing of peptides to disrupt PPIs.

ModLink + : prediction of protein structure using protein–protein interactions

The gap between experimentally determined protein sequences and structures is very large and is increasing. To bridge this gap, a range of computational tools have been devised to model the structure of proteins, including for instance the Protein Model Portal [54], i-TASSER [55] or Phyre2 [56]. ModLink + [29] is one of such tools with the particularity that combines advanced remote homology detection methods and known or inferred PPIs. Thanks to the use of PPIs, ModLink + (last updated for this suite in 2015) improves fold prediction accuracy even for sequences with low identity or similarity with templates. ModLink + has been successfully applied in the modelling of zebrafish nucleotide-binding and oligomerization domains 1 and 2 [57] and three novel psychrophilic enzymes of *Glaciozyma antarctica PI12* [58]. Given its nature, ModLink + represents the transition between the interactome- and structure-based tools provided in InteractoMIX.

VORFFIP and M-VORFFIP: prediction of functional sites in proteins

VORFFIP (VP) [30] and Multi-VORRFIP (MVP) [31] are two computational tools designed to predict functional sites and interfaces in proteins evaluating what regions are more likely to interact with other biomolecules such as proteins, peptides, DNA or RNA. Although the prediction of functional sites and interfaces has been approached by other groups (e.g. [59,60]), MVP centralizes the prediction of different interaction types in a single resource while its performance compares favourably with individual, purpose-made, prediction algorithms [31]. Given a protein structure, VP/MVP depicts each residue with a broad range of structural, evolutionary, experimental and energy-based features. Then, this information is integrated into a probabilistic framework by means of a Random Forest classifier and a probabilistic score is computed. Whereas VP [30] predicts only protein interfaces, MVP [31] is trained to further distinguish between peptide-, DNA- and RNA-binding sites.

As an example of practical application, MVP approach has been used to define the region on the Msh4–Msh5 complex more likely to interact with DNA and thus helping the modelling of the quaternary structure of the entire the complex [61]. As usual applications, VP and MVP's predictions can be used to guide the experimental mapping of functional sites by side-directed mutagenesis or as part of V-D^{2}OCK [32], also part of the InteractoMIX suite, a data-guided docking algorithm.

VD^{2}OCK: predicted interface-guided protein docking

Experimental determination of the structure of protein complexes cannot keep pace with ever increasing interactomic data and thus, to close this gap, computational approaches such as protein docking provide useful structural models. Starting from the structures of two unbound proteins docking methods generate a sampling of bound (or docked) conformations, i.e. structural models of the protein complex, followed by the ranking of the models based on a given scoring function. Docking methods can be broadly categorized into two groups: *ab initio* (e.g. ZDOCK [62], HEX [63], GRAMM [64], SwarmDock [65], ClusPro [66]) and data driven (e.g. HADDOCK [67], PatchDock [68], RosettaDock [69]).

V-D^{2}OCK [32] belongs to data-driven protein docking methods. It comprises three major steps: firstly, VORFFIP [30], described above, is used to predict protein-binding sites; secondly, the predicted binding sites are used to guide PatchDock, a docking method based on geometric hashing [68]; finally docking poses are clustered [70] and ranked using three different scoring functions (PathDock [68], ZRANK [71] and ES3DC potential [72]). By using a guided docking, V-D^{2}OCK is fast enough to perform the docking of large set of interactions in a reasonable time. Thus, the method is especially useful for the high-throughput structural docking of genome-wide interactomes. Even though V-D^{2}OCK was developed recently, it participated in the latest CASP–CAPRI competition (manuscript in press).

PCRPi: prediction of critical residues in interfaces and PCRPi-DB: database of annotated hot-spots in protein complexes

The contribution to the binding energy between proteins varies across the residues located at their interface, with small portion of them contributing the most to it, i.e. the

so-called hot-spot of the interaction [33]. The identification and charting of such regions in protein interfaces is central to a number of problems in Biology and Biomedicine [73] and has clear applications both in in drug discovery and protein design (e.g. [74]).

Several computational tools have addressed the prediction of hot spots in interfaces (e.g. Robetta [75], KFC [76] or HotPoint [77]) including PCRPi [34]. As described, the combination of several sources of information is required to better depict the nature of these types of residues [78]. In this sense, PCRPi [34] is based in a machine-learning classifier (Bayesian Network) that integrates physical-, sequence- and structural-based information. PCRPi was extensively benchmarked on two independent sets [79,80] and on the RAS–AntiRAS antibody complex [81] and fares favourably to other resources available in the field [75,76,82]. Common applications include the prediction of potentially important interfaces residues as candidates for site-directed mutagenesis, structure-based protein design (e.g. disrupt or improve interactions) or drug discovery to derive chemical mimics of critical residues in the interface.

An extension of PCRPi is PCRPi-DB, a database of annotated hot-spots in protein complexes [53]. Structural information of PCRPi-DB repository can be browsed, queried and visualized using a bespoken molecular visualizer. PCRPi-DB also aims at understanding of PPIs at interactome-level (i.e. extracting common features within the interactome [83]). PCRPi-DB is weekly updated to include the release of new protein structures in the protein data bank [84].

PiPreD: modelling of orthosteric peptides in protein interfaces

PiPreD [35], is a structure- and knowledge-based, approach to model the conformation of peptides targeting protein interfaces. As a knowledge-based approach, PiPreD relies on a bespoken library of structural motifs derived from interfaces named iMotifs and it uses native structural elements of the targeted interface in the form of disembodied interface residues named anchor residues. The search and sampling of peptides covers the entire interface, ensuring the systematic and comprehensive exploration of the entire interface and an unbiased sampling of the conformations of peptides. PiPreD is therefore a good complement to existing methodologies mainly based in peptide docking or *ab initio* modelling [85–88].

The input of PiPreD is the structure of a protein complex. The modelling of peptides is based on the anchor residues of any cognate partner that interact with the target protein. Upon the structural fitting of iMotifs using an iterative superposition, side chains are grafted using computational design using the Rosetta suite [89] as part of PiPreD.

The natural application of PiPreD is the design of peptides targeting specific interactions. These peptides, when experimentally validated, can be used as leads for drug-design. An example of such is the development of peptides to target RAS [90]. As part of this piece of research, three designed peptides were found to target specifically active RAS and to impede the interaction with a cognate partner and a RAS antibody (manuscript under preparation). Lastly, the pool of sequences generated can be used to design sequence libraries to be used in conjunction with high-throughput peptide synthesis techniques such as peptide microarrays.

Conclusion

In this review we presented a suites of tools designed to address many, multi-scale, levels of PPIs: from the general network complexity, i.e. interactomes, to the detailed atomic description of individual PPI interfaces. These tools have been compiled in a unique, one-stop, online resource: InteractoMIX, The suite provides a unique resource that tackles some of the top interests in current biomedical research in a user-friendly and intuitive way. InteractoMIX is a useful resource for both computational and experimental scientists with the aim to expand their knowledge on a protein networks and pathways, identify key residues related to the function of a particular PPI and, eventually, predict new therapeutic targets and potential interfering peptides that can be used as leads for drugs.

Funding

This work was supported by the Spanish Ministry of Economy (MINECO) and FEDER [grant numbers BIO2014-57518 and BIO2011-22568]; the Generalitat de Catalunya (ACC10) under the European Union TecnioSpring program [grant number TECSPR13-1-0008]; the Research Councils United Kingdom Academic Fellowship scheme (to N.F.F.); the Wellcome Trust ViP Award (to N.F.F.); and the University of Leeds Scholarship (to J.S.).

References

1 Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A. et al. (2001) The sequence of the human genome. Science **291**, 1304 CrossRef PubMed
2 Rolland, T., Taşan, M., Charloteaux, B., Pevzner, S.J., Zhong, Q., Sahni, N., Yi, S., Lemmens, I., Fontanillo, C., Mosca, R. et al. (2014) A proteome-scale map of the human interactome network. Cell **159**, 1212–1226 CrossRef PubMed
3 Oliver, S. (2000) Guilt-by-association goes global. Nature **403**, 601–603 CrossRef PubMed
4 Mullard, A. (2012) Protein–protein interaction inhibitors get into the groove. Nat. Rev. Drug Discov. **11**, 173–175 CrossRef PubMed
5 Haura, E.B. (2012) From modules to medicine: how modular domains and their associated networks can enable personalized medicine. FEBS Lett **586**, 2580–2585 CrossRef PubMed
6 Kerrien, S., Aranda, B., Breuza, L., Bridge, A., Broackes-Carter, F., Chen, C., Duesbury, M., Dumousseau, M., Feuermann, M., Hinz, U. et al. (2012) The IntAct molecular interaction database in 2012. Nucleic Acids Res, 2012 **40**, D841–D846, Database issue CrossRef
7 Xenarios, I., Salwiński, L., Duan, X.J., Higney, P., Kim, S.M. and Eisenberg, D. (2002) DIP, the database of interacting proteins: a research tool for studying cellular networks of protein interactions. Nucleic Acids Res **30**, 303–305 CrossRef PubMed
8 Stark, C., Breitkreutz, B.J., Chatr-Aryamontri, A., Boucher, L., Oughtred, R., Livstone, M.S., Nixon, J., Van Auken, K., Wang, X, Shi, X. et al. (2011) The BioGRID interaction database: 2011 update. Nucleic Acids Res **39**, D698–D704, Database issue CrossRef PubMed

9 Goel, R., Harsha, H.C., Pandey, A. and Prasad, T.S. (2012) Human protein reference database and human proteinpedia as resources for phosphoproteome analysis. Mol. BioSyst. **8**, 453–463 CrossRef PubMed

10 Licata, L., Briganti, L., Peluso, D., Perfetto, L., Iannuccelli, M., Galeota, E., Sacco, F., Palma, A., Nardozza, A.P., Santonico, E. et al. (2012) MINT, the molecular interaction database: 2012 update. Nucleic Acids Res **40**, D857–D861, Database issue CrossRef PubMed

11 Guldener, U., Münsterkötter, M., Oesterheld, M., Pagel, P., Ruepp, A., Mewes, H.W. and Stümpflen, V. (2006) MPact: the MIPS protein interaction resource on yeast. Nucleic Acids Res **34**, D436–D441 CrossRef PubMed

12 Razick, S., Magklaras, G. and Donaldson, I.M. (2008) iRefIndex: a consolidated protein interaction database with provenance. BMC Bioinformatics **9**, 405 CrossRef PubMed

13 Prieto, C. and De Las Rivas, J. (2006) APID: Agile protein interaction DataAnalyzer. Nucleic Acids Res **34**, W298–W302, Web Server issue CrossRef PubMed

14 Garcia-Garcia, J., Guney, E., Aragues, R., Planas-Iglesias, J. and Oliva, B. (2010) Biana: a software framework for compiling biological interactions and analyzing networks. BMC Bioinformatics **11**, 56 CrossRef PubMed

15 UniProt, C. (2015) UniProt: a hub for protein information. Nucleic Acids Res **43**, D204–D212, Database issue CrossRef PubMed

16 Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A. et al. (2016) The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res **44**, D279–D285, D1 CrossRef PubMed

17 Huntley, R.P., Sawford, T., Mutowo-Meullenet, P., Shypitsyna, A., Bonilla, C., Martin, M.J. and O'Donovan, C. (2015) The GOA database: gene Ontology annotation updates for 2015. Nucleic Acids Res. **43**, D1057–D1063, Database issue CrossRef PubMed

18 Guney, E., Garcia-Garcia, J. and Oliva, B. (2014) GUILDify: a web server for phenotypic characterization of genes through biological data integration and network-based prioritization algorithms. Bioinformatics **30**, 1789–1790 CrossRef PubMed

19 Guney, E. and Oliva, B. (2012) Exploiting protein–protein interaction networks for genome-wide disease-gene prioritization. PLoS One **7**, e43557 CrossRef PubMed

20 Engin, H.B., Guney, E., Keskin, O., Oliva, B. and Gursoy, A. (2013) Integrating structure to protein–protein interaction networks that drive metastasis to brain and lung in breast cancer. PLoS One **8**, e81035 CrossRef PubMed

21 Planas-Iglesias, J., Guney, E., García-García, J., Robertson, K.A., Raza, S., Freeman, T.C., Ghazal, P. and Oliva, B. (2012) Extending signaling pathways with protein-interaction networks. Application to apoptosis. OMICS **16**, 245–256

22 Garcia-Garcia, J., Schleker, S., Klein-Seetharaman, J. and Oliva, B. (2012) BIPS: BIANA interolog prediction server. A tool for protein–protein interaction inference. Nucleic Acids Res. **40**, W147–W151, Web Server issue CrossRef PubMed

23 Schleker, S., Garcia-Garcia, J., Klein-Seetharaman, J. and Oliva, B. (2012) Prediction and comparison of *Salmonella*–human and *Salmonella–Arabidopsis* interactomes. Chem. Biodivers **9**, 991–1018 CrossRef PubMed

24 Huo, T., Liu, W., Guo, Y., Yang, C., Lin, J. and Rao, Z. (2015) Prediction of host–pathogen protein interactions between *Mycobacterium tuberculosis* and *Homo sapiens* using sequence motifs. BMC Bioinformatics **16**, 100 CrossRef PubMed

25 Planas-Iglesias, J., Marin-Lopez, M.A., Bonet, J., Garcia-Garcia, J. and Oliva, B. (2013) iLoops: a protein–protein interaction prediction server based on structural features. Bioinformatics **29**, 2360–2362 CrossRef PubMed

26 Planas-Iglesias, J., Bonet, J., García-García, J., Marín-López, M.A., Feliu, E. and Oliva, B. (2013) Understanding protein–protein interactions using local structural features. J. Mol. Biol. **425**, 1210–1224 CrossRef PubMed

27 Muratcioglu, S., Presman, D.M., Pooley, J.R., Grøntved, L., Hager, G.L., Nussinov, R., Keskin, O. and Gursoy, A. (2015) Structural modeling of GR interactions with the SWI/SNF chromatin remodeling complex and C/EBP. Biophys. J. **109**, 1227–1239

28 Planas-Iglesias, J., Dwarakanath, H., Mohammadyani, D., Yanamala, N., Kagan, V.E. and Klein-Seetharaman, J. (2015) Cardiolipin interactions with proteins. Biophys. J. **109**, 1282–1294 CrossRef PubMed

29 Fornes, O., Aragues, R., Espadaler, J., Marti-Renom, M.A., Sali, A. and Oliva, B. (2009) ModLink+: improving fold recognition by using protein–protein interactions. Bioinformatics, **25**, 1506–1512 CrossRef

30 Segura, J., Jones, P.F. and Fernandez-Fuentes, N. (2011) Improving the prediction of protein binding sites by combining heterogeneous data and Voronoi diagrams. BMC Bioinformatics **12**, 352 CrossRef PubMed

31 Segura, J., Jones, P.F. and Fernandez-Fuentes, N. (2012) A holistic *in silico* approach to predict functional sites in protein structures. Bioinformatics **28**, 1845–1850 CrossRef PubMed

32 Segura, J., Marín-López, M.A., Jones, P.F., Oliva, B. and Fernandez-Fuentes, N. (2015) VORFFIP-driven dock: V-D2OCK, a fast and accurate protein docking strategy. PLoS One **10**, e0118107 CrossRef PubMed

33 Clackson, T. and Wells, J.A. (1995) A hot spot of binding energy in a hormone–receptor interface. Science **267**, 383–386 CrossRef PubMed

34 Assi, S.A., Tanaka, T., Rabbitts, T.H. and Fernandez-Fuentes, N. (2010) PCRPi: presaging critical residues in protein interfaces, a new computational tool to chart hot spots in protein interfaces. Nucleic Acids Res **38**, e86 CrossRef PubMed

35 Oliva, B. and Fernandez-Fuentes, N. (2015) Knowledge-based modeling of peptides at protein interfaces: PiPreD. Bioinformatics **31**, 1405–1410 CrossRef PubMed

36 Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res **13**, 2498–2504 CrossRef PubMed

37 Yu, H., Luscombe, N.M., Lu, H.X., Zhu, X., Xia, Y., Han, J.D., Bertin, N., Chung, S., Vidal, M. and Gerstein, M. (2004) Annotation transfer between genomes: protein–protein interologs and protein–DNA regulogs. Genome Res **14**, 1107–1118 CrossRef PubMed

38 Apweiler, R., Bairoch, A., Wu, C.H., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M. et al. (2004) UniProt: the universal protein knowledgebase. Nucleic Acids Res **32**, D115, Database issue CrossRef PubMed

39 Murakami, Y. and Mizuguchi, K. (2014) Homology-based prediction of interactions between proteins using averaged one-dependence estimators. BMC Bioinformatics **15**, 213 CrossRef PubMed

40 Kim, E., Kim, H. and Lee, I. (2013) JiffyNet: a web-based instant protein network modeler for newly sequenced species. Nucleic Acids Res **41**, W192–W197, Web Server issue CrossRef PubMed

41 Na, I., Reddy, K.D., Breydo, L., Xue, B. and Uversky, V.N. (2014) A putative role of the Sup35p C-terminal domain in the cytoskeleton organization during yeast mitosis. Mol. Biosyst. **10**, 925–940 CrossRef PubMed

42 Bonet, J., Planas-Iglesias, J., Garcia-Garcia, J., Marín-López, M.A., Fernandez-Fuentes, N. and Oliva, B. (2014) ArchDB 2014: structural classification of loops in proteins. Nucleic Acids Res **42**, D315–D319 CrossRef PubMed

43 Tuncbag, N., Gursoy, A., Nussinov, R. and Keskin, O. (2011) Predicting protein–protein interactions on a proteome scale by matching evolutionary and structural similarities at interfaces using PRISM. Nat. Protocols. **6**, 1341–1354 CrossRef

44 Mosca, R., Ceol, A. and Aloy, P. (2013) Interactome3D: adding structural details to protein networks. Nat. Methods **10**, 47–53 CrossRef PubMed

45 Kacprowski, T., Doncheva, N.T. and Albrecht, M. (2013) NetworkPrioritizer: a versatile tool for network-based prioritization of candidate disease genes or other molecules. Bioinformatics **29**, 1471–1473 CrossRef PubMed

46 Gottlieb, A., Magger, O., Berman, I., Ruppin, E. and Sharan, R. (2011) PRINCIPLE: a tool for associating genes with diseases via network propagation. Bioinformatics **27**, 3325–3326 CrossRef PubMed

47 Santana-Codina, N., Carretero, R., Sanz-Pamplona, R., Cabrera, T., Guney, E., Oliva, B., Clezardin, P., Olarte, O.E., Loza-Alvarez, P., Méndez-Lucas, A. et al. (2013) A transcriptome-proteome integrated network identifies ERp57 as a hub that mediates bone metastasis. Mol. Cell. Proteomics **12**, 2111–2125 CrossRef PubMed

48 Zuberi, K., Franz, M., Rodriguez, H., Montojo, J., Lopes, C.T., Bader, G.D. and Morris, Q. (2013) GeneMANIA prediction server 2013 update. Nucleic Acids Res. **41**, W115–W122, Web Server issue CrossRef PubMed

49 Ghiassian, S.D., Menche, J. and Barabasi, A.L. (2015) A DIseAse MOdule Detection (DIAMOnD) algorithm derived from a systematic analysis of connectivity patterns of disease proteins in the human interactome. PLoS Comput. Biol. **11**, e1004120 CrossRef PubMed

50 Dong, C., Wei, P., Jian, X., Gibbs, R., Boerwinkle, E., Wang, K. and Liu, X. (2015) Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. Hum. Mol. Genet. **24**, 2125–2137 CrossRef PubMed

51 Guney, E. and Oliva, B. (2014) Analysis of the robustness of network-based disease-gene prioritization methods reveals redundancy in the human interactome and functional diversity of disease-genes. PLoS One **9**, e94686 CrossRef PubMed

52 Segura Mora, J., Assi, S.A. and Fernandez-Fuentes, N. (2010) Presaging critical residues in protein interfaces-web server (PCRPi-W): a web server to chart hot spots in protein interfaces. PLoS One **5**, e12352 CrossRef PubMed

53 Segura, J. and Fernandez-Fuentes, N. (2011) PCRPi-DB: a database of computationally annotated hot spots in protein interfaces. Nucleic Acids Res **39**, D755–D760, Database issue CrossRef PubMed

54 Haas, J., Roth, S., Arnold, K., Kiefer, F., Schmidt, T., Bordoli, L. and Schwede, T. (2013), The protein model portal – a comprehensive resource for protein structure and model information. Database (Oxford) **2013**, bat031

55 Yang, J. and Zhang, Y. (2015) I-TASSER server: new development for protein structure and function predictions. Nucleic Acids Res. **43**, W174–181

56 Kelley, L.A. and Sternberg, M.J. (2009) Protein structure prediction on the Web: a case study using the Phyre server. Nat. Protoc. **4**, 363–371 CrossRef PubMed

57 Maharana, J., Sahoo, B.R., Bej, A., Patra, M.C., Dehury, B., Bhoi, G.K., Lenka, S.K., Sahoo, J.R., Rout, A.K. and Behera, B.K. (2014) Structural and functional investigation of zebrafish (Danio rerio) NOD1 leucine rich repeat domain and its interaction with iE-DAP. Mol. Biosyst. **10**, 2942–2953 CrossRef PubMed

58 Parvizpour, S., Razmara, J., Ramli, A.N., Md Illias, R. and Shamsir, M.S. (2014) Structural and functional analysis of a novel psychrophilic beta-mannanase from *Glaciozyma antarctica* PI12. J. Comput. Aided Mol. Des. **28**, 685–698 CrossRef PubMed

59 Sikic, M., Tomic, S. and Vlahovicek, K. (2009) Prediction of protein–protein interaction sites in sequences and 3D structures by random forests. PLoS Comput. Biol. **5**, e1000278 CrossRef PubMed

60 Xiong, Y., Liu, J. and Wei, D.Q. (2011) An accurate feature-based method for identifying DNA-binding residues on protein surfaces. Proteins **79**, 509–517 CrossRef PubMed

61 Rakshambikai, R., Srinivasan, N. and Nishant, K.T. (2013) Structural insights into *Saccharomyces cerevisiae* Msh4–Msh5 complex function using homology modeling. PLoS One **8**, e78753 CrossRef PubMed

62 Pierce, B.G., Hourai, Y. and Weng, Z.P. (2011) Accelerating protein docking in ZDOCK using an advanced 3D convolution library. PLoS One **6**, e24657 CrossRef PubMed

63 Ritchie, D.W. and Venkatraman, V. (2010) Ultra-fast FFT protein docking on graphics processors. Bioinformatics **26**, 2398–2405 CrossRef PubMed

64 Tovchigrechko, A. and Vakser, I.A. (2006) GRAMM-X public web server for protein–protein docking. Nucleic Acids Res **34**, W310–W314, Web Server issue CrossRef PubMed

65 Torchala, M., Moal, I.H., Chaleil, R.A., Fernandez-Recio, J. and Bates, P.A. (2013) SwarmDock: a server for flexible protein–protein docking. Bioinformatics **29**, 807–809 CrossRef PubMed

66 Comeau, S.R., Gatchell, D.W., Vajda, S. and Camacho, C.J. (2004) ClusPro: a fully automated algorithm for protein–protein docking. Nucleic Acids Res **32**, W96–W99, Web Server issue CrossRef PubMed

67 Dominguez, C., Boelens, R. and Bonvin, A.M.J.J. (2003) HADDOCK: a protein–protein docking approach based on biochemical or biophysical information. J. Am. Chem. Soc. **125**, 1731–1737 CrossRef PubMed

68 Schneidman-Duhovny, D., Inbar, Y., Nussinov, R. and Wolfson, H.J. (2005) PatchDock and SymmDock: servers for rigid and symmetric docking. Nucleic Acids Res **33**, W363–W367, Web Server issue CrossRef PubMed

69 Lyskov, S. and Gray, J.J. (2008) The RosettaDock server for local protein–protein docking. Nucleic Acids Res **36**, W233–W238, Web Server issue CrossRef PubMed

70 Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A.E. and Berendsen, H.J. (2005) GROMACS: fast, flexible, and free. J. Comput. Chem. **26**, 1701–1718 CrossRef PubMed

71 Pierce, B. and Weng, Z. (2007) ZRANK: reranking protein docking predictions with an optimized energy function. Proteins **67**, 1078–1086 CrossRef PubMed

72 Feliu, E., Aloy, P. and Oliva, B. (2011) On the analysis of protein–protein interactions via knowledge-based potentials for the prediction of protein–protein docking. Protein Sci. **20**, 529–541 CrossRef PubMed

73 Cukuroglu, E., Engin, H.B., Gursoy, A. and Keskin, O. (2014) Hot spots in protein–protein interfaces: towards drug discovery. Prog. Biophys. Mol. Biol. **116**, 165–173, 2–3 CrossRef PubMed

74 Patgiri, A., Yadav, K.K., Arora, P.S. and Bar-Sagi, D. (2011) An orthosteric inhibitor of the Ras–Sos interaction. Nat. Chem. Biol. **7**, 585–587 CrossRef PubMed

75 Kortemme, T. and Baker, D. (2002) A simple physical model for binding energy hot spots in protein–protein complexes. Proc. Natl. Acad. Sci. U.S.A. **99**, 14116–14121 CrossRef PubMed

76 Darnell, S.J., LeGault, L. and Mitchell, J.C. (2008) KFC Server: interactive forecasting of protein interaction hot spots. Nucleic Acids Res **36**, W265–W269, Web Server issue CrossRef PubMed

77 Tuncbag, N., Keskin, O. and Gursoy, A. (2010) HotPoint: hot spot prediction server for protein interfaces. Nucleic Acids Res **38**, W402–W406, Web Server issue CrossRef PubMed

78 DeLano, W.L. (2002) Unraveling hot spots in binding interfaces: progress and challenges. Curr. Opin. Struct. Biol. **12**, 14–20 CrossRef PubMed

79 Thorn, K.S. and Bogan, A.A. (2001) ASEdb: a database of alanine mutations and their effects on the free energy of binding in protein interactions. Bioinformatics **17**, 284–285 CrossRef PubMed

80 Fischer, T.B., Arunachalam, K.V., Bailey, D., Mangual, V., Bakhru, S., Russo, R., Huang, D., Paczkowski, M., Lalchandani, V., Ramachandra, C. et al. (2003) The binding interface database (BID): a compilation of amino acid hot spots in protein interfaces. Bioinformatics **19**, 1453–1454 CrossRef PubMed

81 Tanaka, T., Williams, R.L. and Rabbitts, T.H. (2007) Tumour prevention by a single antibody domain targeting the interaction of signal transduction proteins with RAS. EMBO J **26**, 3250–3259 CrossRef PubMed

82 Tuncbag, N., Gursoy, A. and Keskin, O. (2009) Identification of computational hot spots in protein interfaces: combining solvent accessibility and inter-residue potentials improves the accuracy. Bioinformatics **25**, 1513–1520 CrossRef PubMed

83 Thangudu, R.R., Bryant, S.H., Panchenko, A.R. and Madej, T. (2012) Modulating protein–protein interactions with small molecules: the importance of binding hotspots. J. Mol. Biol. **415**, 443–453 CrossRef PubMed

84 Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000) The protein data bank. Nucleic Acids Res. **28**, 235–242 CrossRef PubMed

85 Moon, J.B. and Howe, W.J. (1991) Computer design of bioactive molecules: a method for receptor-based *de novo* ligand design. Proteins **11**, 314 CrossRef PubMed

86 Frenkel, D., Clark, D.E., Li, J., Murray, C.W., RObson, B., Waszkowycz, B. and Westhead, D.R. (1995) PRO_LIGAND: an approach to *de novo* molecular design. 4. Application to the design of peptides. J. Comput. Aided Mol. Des. **9**, 213–225 CrossRef PubMed

87 Roberts, K.E., Cushing, P.R., Boisguerin, P., Madden, D.R. and Donald, B.R. (2012) Computational design of a PDZ domain peptide inhibitor that rescues CFTR activity. PLoS Comput. Biol. **8**, e1002477 CrossRef PubMed

88 Donsky, E. and Wolfson, H.J. (2011) PepCrawler: a fast RRT-based algorithm for high-resolution refinement and binding affinity estimation of peptide inhibitors. Bioinformatics **27**, 2836–2842 CrossRef PubMed

89 Leaver-Fay, A., Tyka, M., Lewis, S.M., Lange, O.F., Thompson, J., Jacak, R., Kaufman, K., Renfrew, P.D., Smith, C.A., Sheffler, W. et al. (2011) ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. Methods Enzymol **487**, 545–574 CrossRef PubMed

90 Milburn, M.V., Tong, L., deVos, A.M., Brünger, A., Yamaizumi, Z., Nishimura, S. and Kim, S.H. (1990) Molecular switch for signal transduction: structural differences between active and inactive forms of protooncogenic ras proteins. Science **247**, 939–945 CrossRef PubMed

Received 11 December 2015
doi:10.1042/BST20150001