D724–D729 *Nucleic Acids Research, 2011, Vol. 39, Database issue* *Published online 19 November 2010*
doi:10.1093/nar/gkq1200

DIMA 3.0: Domain Interaction Map

Qibin Luo[1], Philipp Pagel[1,2], Baiba Vilne[3] and Dmitrij Frishman[1,2,*]

[1]Department of Genome Oriented Bioinformatics, Technische Universität München, Wissenschaftszentrum Weihenstephan, 85350 Freising, [2]Institute for Bioinformatics and Systems Biology/MIPS, HMGU – German Research Center for Environmental Health, Ingolstädter Landstraße 1, 85764 Neuherberg and [3]Hematology Research Laboratory, Technische Universität München, Trogerstraße 32, 81675 München, Germany

Received September 16, 2010; Revised November 5, 2010; Accepted November 6, 2010

ABSTRACT

Domain Interaction MAp (DIMA, available at http://webclu.bio.wzw.tum.de/dima) is a database of predicted and known interactions between protein domains. It integrates 5807 structurally known interactions imported from the iPfam and 3did databases and 46 900 domain interactions predicted by four computational methods: domain phylogenetic profiling, domain pair exclusion algorithm correlated mutations and domain interaction prediction in a discriminative way. Additionally predictions are filtered to exclude those domain pairs that are reported as non-interacting by the Negatome database. The DIMA Web site allows to calculate domain interaction networks either for a domain of interest or for entire organisms, and to explore them interactively using the Flash-based Cytoscape Web software.

INTRODUCTION

The Domain Interaction MAp (DIMA) is a comprehensive database of domain–domain interactions (DDI) (1,2). It integrates experimentally confirmed domain interactions derived from known three-dimensional structures of protein complexes and links between conserved protein domains predicted by a battery of bioinformatics methods. Protein domain networks reconstructed by DIMA represent a useful tool in many areas of biological research, such as cellular signaling.

Over the past decade, several databases and prediction methods specifically addressing DDIs, as opposed to protein–protein interactions (PPIs), have been developed. While there are many methods for experimental elucidation of DDIs, protein structure determination by X-ray crystallography or NMR spectroscopy currently remains the main systematic source of such information available from public databases. A number of resources have been proposed in which DDIs are inferred from high-resolution three-dimensional structures of interacting proteins [iPfam (3), 3did (4), SCOPPI (5), IBIS (6), PIBASE (7), PSIbase (8), InterPare (9) and SNAPPI (10)]. Since DIMA adopts domain definitions according to the widely popular Pfam database (11), we utilize iPfam and 3did (which are both also based on Pfam) as the source of structure-derived DDIs.

Structural data are highly reliable and help obtain valuable insights into the details of domain interactions, but they are sparse and provide low coverage of DDIs due to substantial difficulties in experimental determination of complex structures. Alternatively, DDIs can be predicted from protein sequences, genomes and high-throughput interaction data.

Phylogenetic profiling of domains was the first computational method integrated in DIMA for constructing domain interaction maps (1,12). This method was inspired by the well-known approach of protein phylogenetic profiling, which was introduced as a means of predicting functional links and physical interactions among proteins by analyzing the presence or absence of orthologs over a large number of genomes (13). The domain pair exclusion algorithm (DPEA) (14) for inferring DDIs from PPI data was integrated into DIMA in 2007 (2).

In the new version of DIMA described here, we added two additional methods to predict DDIs based on entirely different principles. First, we utilize the correlated mutations method in three different variations [McBASC (15), OMES (16), ELSC (17)] to infer DDIs from PPIs obtained from the IntAct database (18). Second, we have integrated the discriminative approach called domain interaction prediction in a discriminative way (DIPD) (19) that predicts DDIs from PPIs and non-PPIs (proteins presumed not to interact) based on machine learning.

To summarize, the new release of DIMA (version 3.0) combines structural information from two different sources with four prediction techniques to derive domain

*To whom correspondence should be addressed. Tel: +49 8161 712134; Fax: +49 8161 712186; Email: d.frishman@wzw.tum.de

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

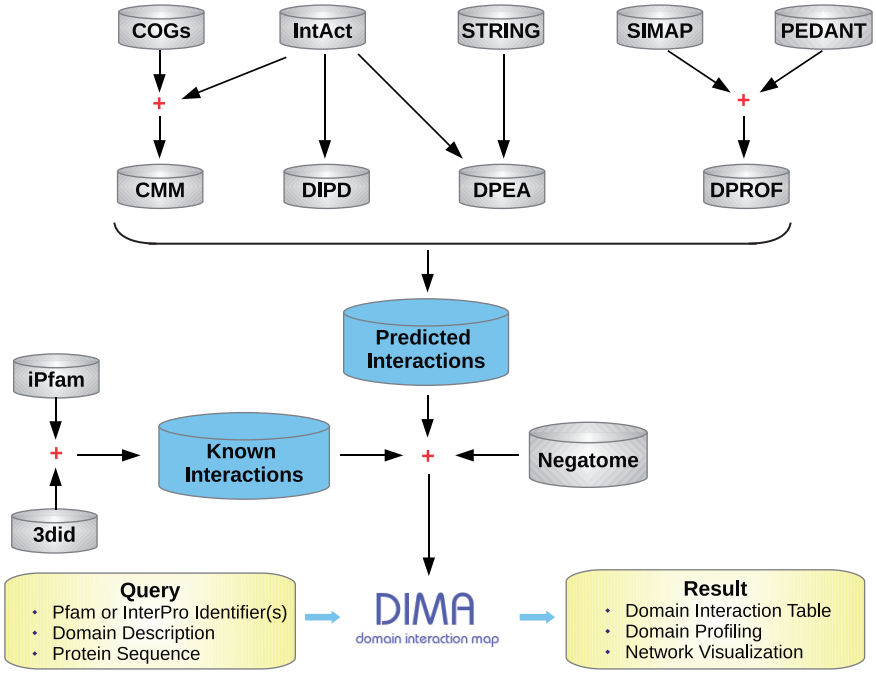

Figure 1. An overview of the DIMA database. Domain interactions are predicted by four computational methods: CMM (correlated mutations), DIPD, DPEA and DPROF (domain phylogenetic profiling). Arrows indicate that a data set or a query is passed to a method or stored as a new data set. Some data sets are combined in a new data set represented using the plus symbol.

interaction maps. An overview of DIMA is shown in Figure 1. The set of features offered by DIMA differs from other integrated resources of this kind. DOMINE (20) integrates DDIs predicted by eight computational methods by importing the original data provided in the respective publications. In contrast, we focus on re-computing all predictions based on current input data. This allows for more up-to-date predictions, including proteins and domains not present in the original reports, and for easy adjustment of method preferences. Re-computation also ensures that all prediction algorithms are run on a common set of input domains. Additionally, DIMA is now updated every 4 months and offers dynamic network visualization. InterDom (21) utilizes fewer methods than DIMA and has not been updated since 2007. Neither DOMINE nor InterDom offer network visualization.

In this study, we describe the content of DIMA 3.0 and report major changes concerning both the computational methods employed to derive domain interactions and the functionality of the Web site.

UPDATE OF DATA SOURCES

Domain definitions

DIMA is based on the domain definitions from Pfam-A (22). As of this writing we are using the Pfam release 24.0, which contains a total of 11 912 protein families (11).

Completely sequenced genomes

Domain phylogenetic profiling involves recording the presence or absence of domains in completely sequenced genomes. We use SIMAP (23) and PEDANT (24) as our source of genomic data and functional annotation. Compared with the previous release of DIMA (2), which contained 460 completely sequenced prokaryotic and eukaryotic genomes, domain profiling is now carried out on 983 complete genomes, almost doubling the coverage.

Structural data

Although DDIs inferred from known structures of protein complexes can only explain 4–19% of the available PPIs (25), this data set can be considered an available gold standard because of its experimental support. The iPfam database was developed by identifying the regions in every protein structure that match a Pfam domain and then generating interacting domain pairs in those cases where the domains are sufficiently close in space (3). An alternative database of domains interacting in 3D is 3did (4). In spite of the very similar approach employed by these resources, only about 66% of DDIs in 3did are confirmed by iPfam. We therefore consider them complementary and import both data sets. The current versions of iPfam and 3did contain 4030 and 5268 unique domain pairs, respectively. We have included these two databases in DIMA, with a union list of 5807 distinct structural DDIs.

High-throughput data on protein interactions

Large-scale experimental data on binary protein interactions obtained by methods such as two-hybrid essay are available from a variety of comprehensive resources. Recently, these databases have formed the International Molecular Exchange consortium (IMEx) and exchange

interaction data regularly. DIMA imports PPIs from the IntAct database (18). The current version of IntAct contains over 200 000 curated binary interaction evidences, from which we inferred 147 722 possible interacting domain pairs in 202 different species.

Predicted protein interactions

In addition to experimental PPI data, we use predicted PPIs from the STRING database (26). STRING predicts functional associations between both individual proteins and orthologous groups of proteins (COGs) (27) using several computational methods, including conserved gene order, phylogenetic profiling, gene fusion and literature mining. DIMA utilizes a high-confidence subset of STRING predictions for COGs (using a conservative threshold of 0.9) as input for subsequent DPEA analysis. In the new version of DIMA, we derived 12 288 DDIs from 118 537 COG interactions.

Data on non-interacting proteins

In DIMA 3.0 we also integrate the Negatome database that contains information on unlikely physical interactions at the protein and domain level (28). A list of non-interacting domain pairs was extracted from Negatome and used to filter all DDIs generated by different computational methods. The current version of Negatome contains 979 unique non-interacting domain pairs. There are 29 and 10 DDIs filtered out by Negatome in 3did and iPfam, respectively. The number of invalidated DDIs in computational methods are as follows: 145 in domain phylogenetic profiling, 1 in correlated mutations, 65 in DPEA for IntAct, 530 in DPEA for STRING and 10 in DIPD.

NEWLY INTEGRATED METHODS

Correlated mutations

The fundamental idea behind the correlated mutation method is functional constraints at the molecular level, namely, the evidence of co-evolution between interacting proteins as well as intra-protein residue pairs. Based on this idea, the correlated mutation method can also be applied at the domain level, where the concept of co-evolution can be extended to domain residue pairs enabling the interactions of domain pairs. The underlying idea is that interacting domains should co-evolve to maintain structural and functional complementarity and that co-evolution of interacting domains can be detected by the presence of compensatory substitutions in the corresponding domain sequences of distinct organisms. Co-evolution between protein domains has been thoroughly documented (29).

A reference set of PPIs is extracted from the IntAct database, ignoring inter-species (e.g. parasite–host) interactions. To guarantee that the co-variation signal corresponds to inter-protein interactions, only hetero-protein pairs are considered.

For each protein from the reference PPI data set, its orthologs are obtained from the STRING database (26,30). Multiple sequence alignment (MSA) of each orthologous group is carried out using the version 3.7 of MUSCLE (31) with all default parameters. We are not using pre-computed Pfam alignments because for our analysis we need to exclude paralogs as well as nearly identical sequences in order to avoid undue bias. As the correlated mutation analysis is subject to several constraints on the size and diversity of the interacting protein families (data not shown), it is carried out only for those PPIs that meet all the following demands: (i) the pairwise sequence alignment length between the reference protein and each ortholog covers at least 80% of each sequence's length; (ii) each species is represented only by one protein sequence in the protein family; (iii) the pairwise sequence identity between all the sequences in the family is below 90%; and (iv) both interacting partners have orthologs in at least a common set of 30 species. We tested different threshold values for the number of species reported previously, from as few as 10 (32) to 100 (33) and found no substantial difference in the quality of results although, of course, more relaxed cutoffs will yield a higher number of predictions (34).

We apply three algorithms that have previously been shown to be among the best performing correlated mutation detection algorithms (33): (i) McBASC (15) calculates correlation coefficient between each pair of sequence positions; (ii) OMES (Observed Minus Expected Squared) (16,35) utilizes a variation of the χ^2 goodness-of-fit test to calculate the statistical significance of the difference between observed and expected frequencies; and (iii) ELSC (Explicit Likelihood of Subset Covariation) (17) is based on the perturbation of a MSA.

All combinatorially possible DDIs in the given reference PPI set are considered and correlation scores among all residues for each domain pair are obtained by extracting those MSA regions that corresponds to the respective domains. For each putative DDI, pairs of positions are sorted by their correlation score and the best scoring pair is defined as predicted contact. In particular, since the OMES algorithm is based on the χ^2 goodness-of-fit test, we further calculate the *P*-values and then combine the obtained *P*-values using the Fisher's combined probability test to get the combined score for each putative DDI. To assess the performance of predictions based on correlated mutations, we benchmarked these three methods against a common reference set of structural domain interactions from iPfam and 3did and found the performance to be similar in McBASC and OMES algorithms, with the ELSC algorithm being best in terms of precision. There are 6968 new interactions predicted by three correlated mutation algorithms (McBASC, OMES, ELSC) with all default parameters, of which 61 interactions are confirmed by iPfam or 3did.

DIPD

Many methods that predict DDIs based on statistical analysis of PPI data and domain composition of the interaction partners do not explicitly take non-PPIs into account. Instead, they rely on enrichment of relevant

features in interacting entities as compared with the entire proteome. However, some studies have proven that including available data on non-PPIs can improve the quality of DDI prediction (14,19). In DIMA 3.0, we integrated a discriminative approach called DIPD (19) for predicting DDIs from PPIs. This approach utilizes both PPIs and non-PPIs to construct domain combinations and then formulates DDI prediction as a feature selection problem in machine learning.

For the DIPD method; the PPI and non-PPI data sets are constructed based on the IntAct database. We generate non-interacting protein pairs randomly from the PPI data set and then exclude: (i) known PPIs, (ii) PPIs whose both interacting partners do not belong to the same species, (iii) PPIs which do not contain any possible combination of two domains found in known PPIs; and (iv) all protein pairs whose orthologs are known to interact. All possible combinations of two domains derived from PPIs are then treated as features. Subsequently, a minimum set of informative features that discriminate PPIs from non-PPIs is delineated using the DIPD method.

SCORING

As described above DIMA 3.0 incorporates four computational methods, each using a different data source. The domain phylogenetic profiling method requires genomic information, DPEA infers DDIs from known PPIs, correlated mutations are detected in protein sequence alignments and the DIPD method operates with PPI and non-PPI data. As these methods are based on different types of data, paradigms and statistics they produce very different scores that are not directly comparable. In addition to combining all the predicted scores in a final combined score, we compute a compressed score for each predicted domain pair and for each computational method as (raw score − min score)/(max score − min score), where the raw score is the original predicted score from a given method, the min score is the lowest score and the max score is the highest score among the predicted scores in a given method. Such compressed scores help to compare different approaches and allow the user to better understand the preferences in the DIMA database.

Obviously, the best approach would be to calibrate each score against known domain interactions and compute precision or accuracy. However, validation and benchmarking of domain prediction methods is not easily accomplished because the only gold standard source of DDIs are databases that are very small when compared with the number of predicted DDIs. Nevertheless, we compared the performance among the methods integrated in DIMA based on structural domain interactions derived from iPfam and 3did. Preliminary results show that DPEA and DIPD produce the most reliable results compared with the other methods with respect to precision. A detailed benchmark study of integrated DIMA predictions will be presented elsewhere (Q. Luo *et al.*, in preparation).

NEW FEATURES OF THE WEB INTERFACE

Network visualization

Previous versions of DIMA offered an option to display a static graphical representation of a domain subgraph with limited possibilities for user interaction. DIMA 3.0 uses the Cytoscape Web API (36) for visualizing and manipulating graphs of DDIs. This interactive network visualization tool models the popular Cytoscape software, but uses the Flash technology rather than Java to reduce launch time; it is compatible with any Web browser. The available version of Cytoscape Web works best with up to a few hundred nodes and edges. The DDI network is passed to the Cytoscape Web API with appropriate parameters, resulting in a dynamic display of graphs that enables users to move nodes and obtain edge information. Additionally, the network can be panned and zoomed in the same layout, edges can be colored according to the method used to predict a particular interaction (e.g. iPfam interactions; green, DPEA interactions; yellow, etc.) and the edge width can be manipulated to represent the interaction score.

Website architecture

The new version of DIMA has been designed to be more easily extensible and maintainable. It has been re-structured for better usability and offers extensive help. The website is built based on the JSP-Model-View-Controller method and uses AJAX technology ('Asynchronous JavaScript and XML') to transport the requested information. The new integrated methods were implemented by using Python.

The web interface allows users to search domain interactions by single or multiple domain identifiers, domain description or sequence. As shown in Figure 2, DIMA results are not only presented as a concise table, but are also displayed using a dynamic graphical representation of the local domain neighborhood. The domain phylogenetic profiling results for a query can be directly accessed in a separate tab.

The entire DDI network can be visualized interactively or obtained by email. Users can easily change a variety of parameters such as distance metrics and thresholds for domain phylogenetic profiling, DPEA cutoffs, thresholds and organism set for the correlated mutation method, etc. The website offers links to all external sources used by the system. Intermediate data, such as tables of phylogenetic profiles, are available for download.

ACKNOWLEDGEMENTS

The authors wish to thank Thomas Rattei, Patrick Tischler and Roland Arnold for their help with the SIMAP resource and Xing-Ming Zhao for assisting us with the DIPD approach. We are grateful to Angelika Fuchs for sharing with us her software to calculate correlated mutations and to Andreas Kirschner and Jan Kirrbach for many helpful discussions and suggestions.

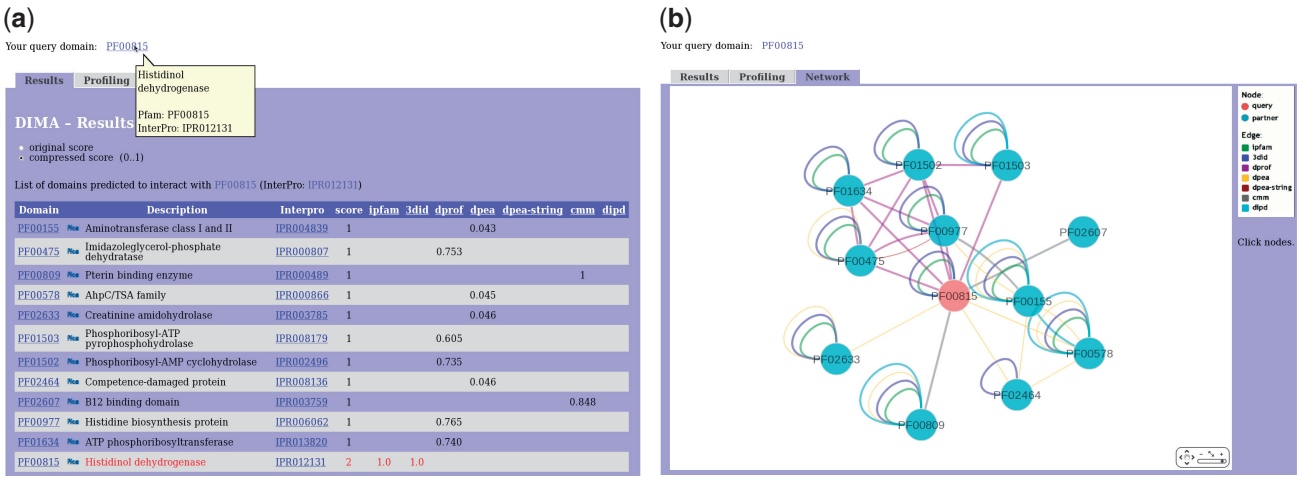

Figure 2. (a) DIMA results are presented in tabular form. The table provides information on interacting partners, their descriptions, InterPro associations and scores. **(b)** Dynamic graphical representation of a domain interaction network by Cytoscape Web. Pfam domains are shown as blue circles. Edges represent predicted or known interactions and are colored according to computational methods and data sources used (e.g. iPfam interactions; green, DPEA interactions; yellow, etc.). The edge width represents the interaction score. The query node(s) are shown in red.

FUNDING

Q.L. is supported by a scholarship of the German Academic Exchange Service (DAAD). Funding for open access charge: Helmholtz Gesellschaft.

Conflict of interest statement. None declared.

REFERENCES

1. Pagel,P., Oesterheld,M., Stümpflen,V. and Frishman,D. (2006) The DIMA web resource–exploring the protein domain network. *Bioinformatics*, **22**, 997–998.
2. Pagel,P., Oesterheld,M., Tovstukhina,O., Strack,N., Stümpflen,V. and Frishman,D. (2008) DIMA 2.0–predicted and known domain interactions. *Nucleic Acids Res.*, **36**, D651–D655.
3. Finn,R.D., Marshall,M. and Bateman,A. (2005) iPfam: visualization of protein-protein interactions in PDB at domain and amino acid resolutions. *Bioinformatics*, **21**, 410–412.
4. Stein,A., Panjkovich,A. and Aloy,P. (2009) 3did Update: domain-domain and peptide-mediated interactions of known 3D structure. *Nucleic Acids Res.*, **37**, D300–D304.
5. Winter,C., Henschel,A., Kim,W.K. and Schroeder,M. (2006) SCOPPI: a structural classification of protein-protein interfaces. *Nucleic Acids Res.*, **34**, D310–D314.
6. Shoemaker,B.A., Zhang,D., Thangudu,R.R., Tyagi,M., Fong,J.H., Marchler-Bauer,A., Bryant,S.H., Madej,T. and Panchenko,A.R. (2010) Inferred Biomolecular Interaction Server–a web server to analyze and predict protein interacting partners and binding sites. *Nucleic Acids Res.*, **38**, D518–D524.
7. Davis,F.P. and Sali,A. (2010) The overlap of small molecule and protein binding sites within families of protein structures. *PLoS Comput. Biol.*, **6**, e1000668.
8. Gong,S., Yoon,G., Jang,I., Bolser,D., Dafas,P., Schroeder,M., Choi,H., Cho,Y., Han,K., Lee,S. *et al.* (2005) PSIbase: a database of Protein Structural Interactome map (PSIMAP). *Bioinformatics*, **21**, 2541–2543.
9. Gong,S., Park,C., Choi,H., Ko,J., Jang,I., Lee,J., Bolser,D.M., Oh,D., Kim,D.-S. and Bhak,J. (2005) A protein domain interaction interface database: InterPare. *BMC Bioinformatics*, **6**, 207.
10. Jefferson,E.R., Walsh,T.P., Roberts,T.J. and Barton,G.J. (2007) SNAPPI-DB: a database and API of Structures, iNterfaces and Alignments for Protein-Protein Interactions. *Nucleic Acids Res.*, **35**, D580–D589.
11. Finn,R.D., Mistry,J., Tate,J., Coggill,P., Heger,A., Pollington,J.E., Gavin,O.L., Gunasekaran,P., Ceric,G., Forslund,K. *et al.* (2010) The Pfam protein families database. *Nucleic Acids Res.*, **38**, D211–D222.
12. Pagel,P., Wong,P. and Frishman,D. (2004) A domain interaction map based on phylogenetic profiling. *J. Mol. Biol.*, **344**, 1331–1346.
13. Pellegrini,M., Marcotte,E.M., Thompson,M.J., Eisenberg,D. and Yeates,T.O. (1999) Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc. Natl. Acad. Sci. USA*, **96**, 4285–4288.
14. Riley,R., Lee,C., Sabatti,C. and Eisenberg,D. (2005) Inferring protein domain interactions from databases of interacting proteins. *Genome Biol.*, **6**, R89.
15. Olmea,O. and Valencia,A. (1997) Improving contact predictions by the combination of correlated mutations and other sources of sequence information. *Fold Des.*, **2**, S25–S32.
16. Kass,I. and Horovitz,A. (2002) Mapping pathways of allosteric communication in GroEL by analysis of correlated mutations. *Proteins*, **48**, 611–617.
17. Dekker,J.P., Fodor,A., Aldrich,R.W. and Yellen,G. (2004) A perturbation-based method for calculating explicit likelihood of evolutionary co-variance in multiple sequence alignments. *Bioinformatics*, **20**, 1565–1572.
18. Aranda,B., Achuthan,P., Alam-Faruque,Y., Armean,I., Bridge,A., Derow,C., Feuermann,M., Ghanbarian,A.T., Kerrien,S., Khadake,J. *et al.* (2010) The IntAct molecular interaction database in 2010. *Nucleic Acids Res.*, **38**, D525–D531.
19. Zhao,X.-M., Chen,L. and Aihara,K. (2010) A discriminative approach for identifying domain-domain interactions from protein-protein interactions. *Proteins*, **78**, 1243–1253.
20. Raghavachari,B., Tasneem,A., Przytycka,T.M. and Jothi,R. (2008) DOMINE: a database of protein domain interactions. *Nucleic Acids Res.*, **36**, D656–D661.
21. Ng,S.-K., Zhang,Z., Tan,S.-H. and Li,K. (2003) InterDom: a database of putative interacting protein domains for validating predicted protein interactions and complexes. *Nucleic Acids Res.*, **31**, 251–254.
22. Sonnhammer,E.L., Eddy,S.R. and Durbin,R. (1997) Pfam: a comprehensive database of protein domain families based on seed alignments. *Proteins*, **28**, 405–420.
23. Rattei,T., Tischler,P., Götz,S., Jehl,M.-A., Hoser,J., Arnold,R., Conesa,A. and Mewes,H.-W. (2010) SIMAP–a comprehensive database of pre-calculated protein sequence similarities, domains, annotations and clusters. *Nucleic Acids Res.*, **38**, D223–D226.

24. Walter,M.C., Rattei,T., Arnold,R., Güldener,U., Münsterkötter,M., Nenova,K., Kastenmüller,G., Tischler,P., Wölling,A., Volz,A. *et al.* (2009) PEDANT covers all complete RefSeq genomes. *Nucleic Acids Res.*, **37**, D408–D411.
25. Schuster-Böckler,B. and Bateman,A. (2007) Reuse of structural domain-domain interactions in protein networks. *BMC Bioinformatics*, **8**, 259.
26. Jensen,L.J., Kuhn,M., Stark,M., Chaffron,S., Creevey,C., Muller,J., Doerks,T., Julien,P., Roth,A., Simonovic,M. *et al.* (2009) STRING 8–a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res.*, **37**, D412–D416.
27. Tatusov,R.L., Fedorova,N.D., Jackson,J.D., Jacobs,A.R., Kiryutin,B., Koonin,E.V., Krylov,D.M., Mazumder,R., Mekhedov,S.L., Nikolskaya,A.N. *et al.* (2003) The COG database: an updated version includes eukaryotes. *BMC Bioinformatics*, **4**, 41.
28. Smialowski,P., Pagel,P., Wong,P., Brauner,B., Dunger,I., Fobo,G., Frishman,G., Montrone,C., Rattei,T., Frishman,D. *et al.* (2010) The Negatome database: a reference set of non-interacting protein pairs. *Nucleic Acids Res.*, **38**, D540–D544.
29. Yeang,C.-H. and Haussler,D. (2007) Detecting coevolution in and among protein domains. *PLoS Comput. Biol.*, **3**, e211.
30. Muller,J., Szklarczyk,D., Julien,P., Letunic,I., Roth,A., Kuhn,M., Powell,S., von Mering,C., Doerks,T., Jensen,L.J. *et al.* (2010) eggNOG v2.0: extending the evolutionary genealogy of genes with enhanced non-supervised orthologous groups, species and functional annotations. *Nucleic Acids Res.*, **38**, D190–D195.
31. Edgar,R.C. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, **5**, 113.
32. Jothi,R., Cherukuri,P.F., Tasneem,A. and Przytycka,T.M. (2006) Co-evolutionary analysis of domains in interacting proteins reveals insights into domain-domain interactions mediating protein-protein interactions. *J. Mol. Biol.*, **362**, 861–875.
33. Halperin,I., Wolfson,H. and Nussinov,R. (2006) Correlated mutations: advances and limitations. A study on fusion proteins and on the Cohesin-Dockerin families. *Proteins*, **63**, 832–845.
34. Kowarsch,A., Fuchs,A., Frishman,D. and Pagel,P. (2010) Correlated mutations: a hallmark of phenotypic amino acid substitutions. *PLoS Comput. Biol.*, **6**, e1000923.
35. Fodor,A.A. and Aldrich,R.W. (2004) Influence of conservation on calculations of amino acid covariance in multiple sequence alignments. *Proteins*, **56**, 211–221.
36. Lopes,C.T., Franz,M., Kazi,F., Donaldson,S.L., Morris,Q. and Bader,G.D. (2010) Cytoscape Web: an interactive web-based network browser. *Bioinformatics*, **26**, 2347–2348.